DOCUMENT RESUME

ED 032 760 24 EM 007 313

By-Mulac, Anthony John
An Experimental Study of the Relative Pedagogical Effectiveness of Videotape and Audiotape Playback of Student Speeches for Self-Analysis in a Basic Speech Course. Final Report.
Eastern Michigan Univ., Ypsilanti.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-7-E-005
Pub Date Dec 68
Grant-OEG-3-7-070005-1927
Note-234p.
EDRS Price MF-$1.00 HC-$11.80
Descriptors-Audiovisual Aids, Aural Learning, Basic Skills, *Feedback, Language Skills, Oral Communication, Public Speaking, Reinforcement, *Self Evaluation, Speaking Activities, Speech Education, Speech Evaluation, *Speech Improvement, *Speech Instruction, Speech Skills, Tape Recordings, Teaching Techniques, *Video Tape Recordings, Visual Learning
Identifiers-Price Speech Performance Rating Scale

To test the relative pedagogical value of electronic feedback within the Eastern Michigan University's fundamentals of speech program, the following general hypothesis was established: the greater the completeness and accuracy of student speech performance feedback, the greater the degree of speech skill a student sill later exhibit. Feedback was defined as any consequence or result of performance that is perceived by the learner. Behaviorally, three levels of feedback were seen: videotape replay of two class performances with traditional (class and instructor) feedback for all performances; audiotape replay of two class performances with traditional feedback for all performances; no electronic replay of any class performances but with traditional feedback for all performances. Videotapes of three other speakers were viewed. Native speech skill and improvement were measured by nine evaluators who viewed videotapes of the first and final speeches made by all students. Analysis indicated that subjects who viewed videotapes of two of their class performances demonstrated significantly greater overall speech skill and bodily action, personality, language, and voice skills than the other two groups between whom there were no significant statistical differences. [Photographs in Appendix 2 have been deleted due to poor reproductibility]. (MM)

ED032760

FINAL REPORT
Project No. 7-E-005
Grant No. OEG 3-7-070005-1927

AN EXPERIMENTAL STUDY OF THE RELATIVE PEDAGOGICAL EFFECTIVENESS OF VIDEOTAPE AND AUDIOTAPE PLAYBACK OF STUDENT SPEECHES FOR SELF-ANALYSIS IN A BASIC SPEECH COURSE

December 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EM007313

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Final Report
Project No. 7-E-005
Grant No. OEG 3-7-070005-1927

AN EXPERIMENTAL STUDY OF THE RELATIVE PEDAGOGICAL EFFECTIVENESS OF VIDEOTAPE AND AUDIOTAPE PLAYBACK OF STUDENT SPEECHES FOR SELF-ANALYSIS IN A BASIC SPEECH COURSE*

Anthony John Mulac

Eastern Michigan University
Ypsilanti, Michigan 48197

December 1968

The research reported herein was performed pursuant to a grant from the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

*Original Title: Pilot Study of the Effects of Audio-Visual Television Playback of Student Speeches for Self-Analysis in a Basic Speech Course

PREFACE

A Brief Summary of This Study

It was the purpose of this study to test the relative pedagogical value of three conditions of electronic feedback within the university fundamentals of speech program. To this end, a general hypothesis was established on the basis of contemporary learning theory: "The greater the completeness and accuracy of student speech performance feedback, the greater the degree of speech skill a student will later exhibit."

Feedback, the independent variable of greatest importance in testing this hypothesis, was defined as "any consequence or results of performance that are perceived by the learner." It was operationalized within the context of a university fundamentals of speech course as three levels of feedback completeness: (1) videotape replay of two class performances plus traditional (class and instructor) feedback for all performances; (2) audiotape replay of two class performances plus traditional feedback for all performances; (3) no electronic replay of any class performances but traditional feedback for all performances and viewing videotapes of three other speakers.

The second independent variable was the subjects' two instructors representing dissimilar interests and backgrounds: (1) public address and group communication; (2) theater. Measurement of the two trials for each subject represented the third independent variable: (1) pretest performance; (2) post-test performance. These three levels of feedback, two instructors, with two trials per subject represented a 3 X 2 X 2 "fixed effects" model.

Subjects were 108 Eastern Michigan University students enrolled in seven sections of Fundamentals of Speech during the Spring Semester, 1967. They were stratified on age, sex, speech background, and instructor, and randomly assigned to the three feedback treatment groups. Each of the seven sections contained approximately equal numbers of subjects assigned to each feedback condition.

To measure the dependent variable, speech skill, five faculty and four student judges were trained in the use of an eleven-point version of the Price multi-factor speech performance rating scale. Videotape recordings of the subjects' first and final speeches were stratified on the three levels of feedback, two instructors, and two trials, and randomly ordered in groups of twelve for evaluation. Judging of these 216 speeches was accomplished in ten rating sessions with an overall reliability of .95 estimated by the Ebel intra-class procedure.

Pretest and post-test performance ratings by the judges were subjected to a Winer "Case II" three-way analysis of variance. Planned comparison t tests were employed in place of F tests where the theoretical model predicted particular relationships.

Results of these analyses indicated that subjects who viewed videotapes of two of their class performances demonstrated significantly greater overall speech skill ($p < .01$, df = 102) in their final speeches than subjects who heard audiotapes of two of their class performances. This represented a forty per cent greater gain on total speech scores by videotape subjects than by audiotape subjects. In final speeches, members of the videotape group were also significantly better than the audiotape group on four of the six Price scale factors: Bodily Action, Personality, Language, and Voice.

No statistically significant differences were found between final speeches of subjects who heard audiotape recordings of two of their class performances and subjects who received no electronic feedback but viewed videotapes of three speakers other than themselves. Nor were there significant differences between subjects of the two instructors, nor significant interactions between instructors and treatments.

Subjects without regard to treatment or instructor showed significant improvement ($p < .001$, df = 102) between pretest and post-test performances on overall speech skill and on five of the six Price scale factors: Content, Bodily Action, Personality, Language, and Voice, with greatest improvement on Content.

Within the limits of the subjects and design employed, this study clearly demonstrated the superiority of videotape feedback over audiotape and lack of electronic feedback in the university fundamentals of speech course.

TABLE OF CONTENTS

Page

PREFACE: A BRIEF SUMMARY OF THIS STUDY ii

LIST OF TABLES v

LIST OF APPENDICES viii

LIST OF FIGURES x

LIST OF PICTURES xi

Chapter

I. THE PROBLEM 1

Introduction to the Problem
Hypotheses
Definitions
Theoretical Model
Organization of the Study

II. REVIEW OF RELATED RESEARCH 16

Research Dealing with Pedagogical Uses of Electronic Feedback
Research Dealing with Instruments for Measuring Speech Skill
Summary of Research

III. PRELIMINARY TESTING OF PROCEDURES AND APPARATUS 48

Pilot Test of Videotape Replay Procedure
Pilot Test of the Instrument and Procedure for Measuring Speech Skill
Summary

IV. RESEARCH PROCEDURE 66

Elements of Design
Special Installation of Equipment Required for the Experiment

Chapter Page

Procedure for Employing the Independent Variables
Procedure for Measuring the Dependent Variable
Summary of Procedure

V. RESULTS 93

Reliability of Measurements
Validity of Measurements
Statistical Manipulation of Major Data
Results of Analyses of Major Data
Results of Analyses of Subsidiary Measures
Summary

VI. SUMMARY AND CONCLUSIONS 122

Summary
Conclusions
Future Research
Predictive Generalizations

APPENDICES 136

SELECTED BIBLIOGRAPHY 212

(Appendix J deleted due to poor reproducibility of photographs)

LIST OF TABLES

Table		Page
1.	Ebel Intra-Class Reliability Estimates . .	61
2.	Ebel Estimate of Reliability of Ratings on 216 Speeches	94
3.	Ebel Estimates of Reliability of Ratings on Twelve Speeches with One Judge Omitted	95
4.	Rank Order of Test-Retest Reliability Coefficients for Each Judge on Twelve Speeches	97
5.	Results of Hartley Tests for Homogeneity of Variance of Data	100
6.	Factor Weights for the Price Scale: Two Sources	105
7.	Means and Standard Deviations for Groups Involved in Significant Relationships on Total Weighted$_{\text{Price}}$ Scores	107
8.	Results of 3-Way$_1$ Analysis of Variance on Total Weighted$_{\text{Price}}$ Scores	108
9.	Results of Planned Comparisons on Total Weighted$_{\text{Price}}$ Scores	108
10.	Means and Standard Deviations for Group Scores on Price Scale Factors	111
11.	Summary of Significant Findings of 3-Way$_1$ Analyses of Variance on Factor and Total Weighted Scores	113
12.	Results of 2-Way Analysis of Variance on Subject Gains for Six Factors	113

Table Page

13. Results of Newman-Keuls Procedure on Subject Gain Scores for Each Factor . . 114

14. Results of 3-Way[1] Analysis of Variance on Total Weighted$_{\text{Judges}}$ Scores 201

15. Results of Planned Comparisons on Total Weighted$_{\text{Judges}}$ Scores 201

16. Results of 3-Way[1] Analysis of Variance on Factor #1, "Content," Scores 202

17. Results of Planned Comparisons on Factor #1, "Content," Scores 202

18. Results of 3-Way[1] Analysis of Variance on Factor #2, "Intelligibility," Scores 203

19. Results of Planned Comparisons on Factor #2, "Intelligibility," Scores 203

20. Means and Standard Deviations for Instructor-Repeated Measure Group Interactions on Factor #3, "Bodily Action," Scores 204

21. Results of 3-Way[1] Analysis of Variance on Factor #3, "Bodily Action," Scores . 205

22. Results of Planned Comparisons on Factor #3, "Bodily Action," Scores 205

23. Results of 3-Way[1] Analysis of Variance on Factor #4, "Personality," Scores . . 206

24. Results of Planned Comparisons on Factor #4, "Personality," Scores . . . 206

25. Results of 3-Way[1] Analysis of Variance on Factor #5, "Language," Scores . . . 207

26. Results of Planned Comparisons on Factor #5, "Language," Scores 207

Table Page

27. Results of 3-Way Analysis of Variance on Factor #6,[1] "Voice," Scores 208

28. Results of Planned Comparisons on Factor #6, "Voice," Scores 208

29. Results of 2-Way Analysis of Variance of Subject Pretest Performances on Total Weighted$_{\text{Price}}$ Scores 209

30. Results of 1-Way Analysis of Variance on Subject Total Speech Points Assigned by Instructor #1 210

31. Results of 1-Way Analysis of Variance on Subject Total Speech Points Assigned by Instructor #2 210

32. Results of 1-Way Analysis of Variance on Subject Total Course Points Assigned by Instructor #1 211

33. Results of 1-Way Analysis of Variance on Subject Total Course Points Assigned by Instructor #2 211

LIST OF APPENDICES

Appendix Page

A. Eleven-Step Version of Price Speech Performance Rating Scale 137

B. Instructions for Subjects of Pilot Study on Videotape Recording-Replay Procedure 138

C. Questionnaire Text and Responses for Pilot Test of Videotape Recording-Replay Procedure 140

D. Questionnaire Text and Responses for Pilot Test of Price Scale Judging Procedure 147

E. Text of General Instructions and Rationale Read to All Subjects at First Class Meeting of Major Experiment 151

F. Text of Pretest Questionnaire for Experimental Subjects Given First Day of Class 152

G. Three Different Sets of Instructions for Three Experimental Treatment Groups. . 155

H. Announcement of "Hawthorne Control" Activity 161

I. Text of Instructions and Rating Forms Used in "Hawthorne Control" Activity . 162

J. Pictures of Recording, Playback, and Judging Procedure 167

K. Text and Responses to Post-Test Questionnaire 171

Appendix Page

L. Text and Responses for Major Experiment Post-Judging Questionnaire . . . 196

M. Results of 3-Way$_1$ Analyses of Variance and Planned Comparisons on Major Data 201

N. Results of Analyses of Variance of Subsidiary Data 215

LIST OF FIGURES

Figure Page

1. Pre-Test Scores for Three Treatment Groups on Price Scale Factors 114

2. Post-Test Scores for Three Treatment Groups on Price Scale Factors 114

LIST OF PICTURES

Picture		Page
1.	Speaker at Lectern	167
2.	Camera and Microphone	167
3.	Instructor Conducting Critique Period	167
4.	Television Camera (Masking Board Removed	167
5.	Audio and Video Controls in the Quirk Fan Room	168
6.	Pan-and-Tilt Control Shown with Operator	168
7.	Roosevelt Master Control Room with Engineer and Videotape Recorders	168
8.	Group I (Videotape) Member	169
9.	Viewing Class Performance	169
10.	Group II (Audiotape) Member	169
11.	Listening to Class Performance	169
12.	Judges for Major Experiment	170
13.	Evaluating Experiment Speeches	170

CHAPTER I

THE PROBLEM

Introduction to the Problem

Nearly nineteen hundred years ago, Quintilian summarized the speech teacher's task of providing students with performance feedback in the following words:

> He must control his temper without however shutting his eyes to faults requiring correction. . . . In praising the recitation of his pupils he must be neither grudging nor over-generous: the former quality will give them a distaste for work, while the latter will produce a complacent self—satisfaction. In correcting faults he must avoid sarcasm and above all abuse: for teachers whose rebukes seem to imply positive dislike discourage industry.[1]

Quintilian understood that the line between too much and too little praise is difficult, yet necessary, to draw and maintain.

The mid-twentieth century teacher of speech finds it no easier to solve this problem, a problem which is intensified by the fact that its resolution differs from student to student. However, modern technology has recently provided what some authorities purport to be a partial solution. Videotape recording equipment is capable of preserving a student's performance so that he can observe that performance as it actually occurred and evaluate its quality. Eisenstadt's phrase, "as others see us," perhaps best depicts this unique quality of videotape feedback.[1]

---

[1] Butler, H. E. (trans.), The Institutio Oratoria of Quintilian (Cambridge: Harvard University Press, 1953), I, 213.

Harold E. Nelson suggests the current sentiment among speech teachers when he says, "Videotaping equipment is undoubtedly a teaching aid that lends itself effectively to the speech classroom."[2] Indeed, many of the nation's largest universities, as well as some of its smallest colleges, have utilized videotaping equipment in their program for speech instruction.[3]

This rapid acceptance of videotape replay as a technique for increasing the effectiveness of speech instruction has come in spite of two basic objections to its use: First, purchase of the system requires a large outlay of funds, at least compared with most other teaching aids. A minimal system of videotape recorder, microphone, camera, playback monitor, and ten hours of re-usable recording tape costs between $2,300 and $3,000.[4] Although the system does not require a professional technician for its operation, it does require occasional professional maintenance. Second, and more damaging, is the fact that the pedagogical usefulness of videotape replay has not been demonstrated experimentally. A review of the literature shows that although many studies have dealt with aspects of videotape recording for teaching purposes, none appears to have tested the fundamental question: Does videotape feedback help students achieve greater skill in speaking than they otherwise might acquire in a speech class not employing this technique? The basic purpose of the research reported here is to answer this question within the setting of a university course in fundamentals of speech. The remainder of this chapter will discuss the fundamental hypothesis of this investigation and the theoretical model which supports it.

---

[1] Arthur Eisenstadt, "As Others See Us," Speech Teacher, I (March, 1952), 121-24.

[2] Harold E. Nelson, "Videotaping the Speech Course," Speech Teacher, XVII (March, 1968), 103.

[3] R. V. E. Reynolds, "Videotape in Teaching Speech in a Small College," Speech Teacher, XVII (March, 1968), 113-15.

[4] LaVerne Weber (Coordinator of Instructional Broadcasing, Eastern Michigan University), interview, July 26, 1968.

Hypotheses

The general hypothesis central to this study may be stated as follows: "The greater the completeness and accuracy of student speech performance feedback, the greater the degree of speech skill a student will later exhibit." Put more simply, the more complete the feedback, the better the ensuing speech performance. Two research hypotheses to be tested are based on this general hypothesis:

> Hypothesis: Students who view videotape recordings of two of their own class performances will demonstrate greater speech skill in post-test speeches than students who hear audiotape recordings of two of their class performances.

> Hypothesis: Students who hear audiotape recordings of two of their own class performances will demonstrate greater speech skill in post-test speeches than students who view videotape recordings of three speakers other than themselves.

The third research hypothesis is based on the accepted rationale for university courses in fundamentals of speech:

> Hypothesis: Students will demonstrate greater speech skill on post-test speeches than on pretest speeches, without regard to feedback treatment or instructor.

Finally, questions relating to the impact of the two types of instructor on post-test speeches and the relative reliability of ratings by student and faculty judges will be investigated.

Definitions

For the sake of clarity, theoretical and operational definitions of the more important terms of the general hypothesis are provided below.

Deese and Hulse provide an appropriate theoretical definition of "feedback" in The Psychology of Learning:

> Feedback refers to any consequences or results of performance that are perceived by the learner. It may be in the form of direct perception of the adequacy of performance, or it may be in the form of verbal information about the adequacy of performance. In general, then, feedback occurs either as the sensory consequence of motor action or as knowledge of results supplied by a teacher, an experimenter, or an automatic teaching device.[1]

Cybernetics theorist Norbert Wiener also defines feedback in terms of performance. Although he discusses it in terms of machine performance he makes it clear that he recognizes the application of his description to the effect of feedback upon human performance.

> Feedback is a method of controlling a system by reinserting into it the results of its past performance. If these results are merely used as numerical data for the criticism of the system and its regulation, we have the simple feedback of the control engineers. If, however, the information which precedes backward from the performance is able to change the general method and pattern of performance, we have a process which may well be called learning.[2]

In describing feedback within the context of discussion theory, Barnlund and Haiman also use the analogy of machines:

---

[1] James Deese and Stewart H. Hulse, The Psychology of Learning (3rd ed.; New York: McGraw-Hill, 1967), p. 454.

[2] Norbert Wiener, The Human Use of Human Beings (Garden City, New York: Doubleday and Co., 1956), p. 61.

> Feedback is a term which has been borrowed by social scientists from the physical sciences and refers to that process which is illustrated in the operation of a thermostat. Very simply, the thermometer in the dining room records the temperature of the house and feeds it back to the furnace controls, "telling the furnace" whether it is giving off enough heat or needs to change its "behavior" in order to adapt to the changing climatic conditions.
>
> Similarly, the members of a discussion group can occasionally interrupt their regular business and ask, "How are we doing?" . . . In other words, the group engages in conscious, critical examination of its own interpersonal patterns with a view to determining what, if any, changes are desirable.[1]

These authors place importance upon the use of feedback for the purpose of self-evaluation of past performance in order to change future performance.

Operationally, the term feedback is used in this research project to refer to the playback of a speech student's class performance by means of a videotape or audiotape recording. Since the recording is a "consequence . . . of performance" which is played back so that it may be "perceived by the learner," videotape and audiotape recording replay clearly fit the definition of feedback given earlier. The completeness and accuracy of this feedback is operationally defined by the number of sensory stimuli inherently transmittible by the recording-replay mechanisms employed. That is, since videotape replay recreates both the sight and sound of the performance, it provides "more complete and accurate" feedback than audiotape replay, which in turn provides more than the complete lack of replay of a student's performance. These three levels of feedback operate as an independent variable within the context of this study.

---

[1] Dean C. Barnlund and Franklyn S. Haiman, The Dynamics of Discussion (Boston: Houghton Mifflin, 1960), pp. 204-5.

"Speech skill," defined in broad theoretical terms, is the ability to demonstrate those characteristics of delivery and composition which maximize the probability of orally communicating the intended message to an audience with the intended result. Rather than an inborn trait or an easily acquired knack, speech skill is an ability that comes from knowledge, practice, and aptitude. As taught today, the purpose of speech skill is not to bring aesthetic pleasure to the audience, or to provide a colorful display. Instead its primary function is to aid the speaker in communicating his ideas to an audience so that the members of that audience might be led to understand the facts of a particular situation or event, or led to modify their behavior in some way predetermined by the speaker.

As generally understood in the field of speech, this skill contains several relatively independent dimensions. Although multivariate analysis has led to several different lists of the component factors of speech skill, there seems to be general agreement that there are at least two dimensions of great importance, delivery and composition. Delivery, the speaker's overt behavior during the speech, includes such elements as the speaker's use of voice, bodily action, general poise, and articulation-pronunciation. Composition includes investigation and analysis of the subject, organization of ideas, use of specific material supporting those general points, and choice of language to convey that material.

Operationally, the "degree of speech skill" demonstrated by a subject is directly proportional to the magnitude of his mean ratings assigned by nine trained judges using Price's six-factor speech performance rating scale.[1]

The period of instruction investigated by this study was an entire semester during which the subjects were enrolled in a fundamentals of speech course. Ratings of all subjects' first and final speeches were obtained in order to test the general hypothesis through analysis of variance or pre-treatment and post-treatment scores.

Theoretical Model

The significance of feedback in the acquisition of behavioral skills is supported by Bilodeau and Bilodeau's analysis of experimentation dealing with this concept:

> Studies of feedback or knowledge of results (KR) show it to be the strongest, most important variable controlling performance and learning. It has been shown repeatedly, as well as recently, that there is no improvement without KR, progressive improvement with it, and deterioration after its withdrawal. . . . No other independent variable offers the wide range of possibilities for getting man to repeat, or change his R[esponse]s immediately or slowly, by small or large amounts.[2]

Deese and Hulse accept this view of the impact of feedback upon performance: "There seems to be universal agreement among those who study the learning of skills that the most fundamental condition determining performance during learning is feedback."[1]

---

[1] William K. Price, "The University of Wisconsin Speech Attainment Test" (unpublished Ph.D. dissertation, University of Wisconsin, 1964).

[2] Edward A. Bilodeau and Ina McD. Bilodeau, "Motor-Skills Learning," Annual Review of Psychology, XXII (1961), 250.

In the process of influencing performance, feedback may demonstrate several distinct functions. In 1949, J. S. Brown proposed three separate roles: reward, information, and motivation.[2] In a given situation, according to his theory, knowledge of results might act to reinforce existing habitual behavior, provide cues to evoke already established habits, or provide the incentive to learn new behavior. More recently, Deese and Hulse have combined Brown's three functions of feedback into two: "(1) it informs the learner about his responses and thus permits him to correct these in some way on the next trial, and (2) it may provide reinforcement."[3] These statements regarding the importance and function of feedback in learning and performance suggest theoretical support for the basic hypothesis of this study.

Two general problems related to feedback should be clarified before turning to its application in teaching speech. The first involves the distinction between learning and performance:

> We are led to make a distinction between learning (what the rat knows from its past experience) and performance (what the rat is willing to show us about what it knows at any particular moment in time). In short, learning must always be inferred from overt performance. But sometimes we may have good reason to believe that

---

[1] Deese and Hulse, Learning, p. 454.

[2] J. S. Brown, "A Proposed Program of Research on Psychological Feedback (Knowledge of Results) in the Performance of Psychomotor Tasks," Research Planning Conference on Perceptual and Motor Skills, AFHRRC Conf. Rept. 49-2 (San Antonio: U. S. Air Force, 1949), 81-87.

[3] Deese and Hulse, Learning, p. 454.

> an organism is simply not demonstrating what it has learned.[1]

Feedback can affect "either learning or performance or both."[2] An experiment of Smode offers evidence for this last point, suggesting that feedback directly affects performance, but influences acquisition only to the extent that acquisition is based upon or dependent upon performance.[3] The complex question of operationally differentiating between learning and performance appears to be less important in the current study than in many other studies because in this research the primary concern is the acquisition of performance skill. Measurement of that skill was used to answer the question of whether different conditions of feedback produce different effects.

The second problem involves the question of whether it is reasonable to expect demonstrated skill to be proportional to the completeness of feedback received. Bilodeau and Bilodeau point to a number of studies which have demonstrated that the level of performance following feedback is directly related to the completeness of feedback:

> Studies shortly after World War II used filters to redden the normally white target of a gunnery trainer. The filters operated when S[ubject] was on target and they greatly improved performance. Later, switching to non-filter feedback showed that removing the cue produced an immediate and large letdown.[4]

---

[1] Ibid., p. 62.

[2] Ibid., p. 454.

[3] A. F. Smode, "Learning and Performance in a Tracking Task under Two Levels of Achievement Information Feedback," Journal of Experimental Psychology, LVI (October, 1958), 297-304.

[4] Bilodeau and Bilodeau, "Motor-Skills Learning," p. 257.

Similarly, by varying the completeness of feedback cues, the Smode experiment, cited above, uncovered evidence for the contention that performance is proportional to completeness of feedback.[1] Bilodeau and Bilodeau summarize their analysis of current research on feedback completeness:

> There is, however, no dispute about certain extra K[nowledge of] R[esult]s producing improved responses. Someone can be expected to use this technique to teach his S[ubject]s more about the criterion task.[2]

Research conducted by Harold Nelson demonstrates that different channels of communication convey different amounts of information, presumably because of the complexity of the message inherently transmittible via each.[3] In his experiment, two films dealing with aerodynamics were presented in various manners to eight different groups, each consisting of fifty-four ROTC students. The independent variable, manner of presentation, was tested by presenting the film in various combinations of video and/or audio, on one or both films, with or without lights on. An objective examination demonstrated that the group which saw and heard both films obtained the highest mean score. The lowest score outside the control group was received by the group which was only permitted to hear the two films. However the treatment conditions in Nelson's study are not directly analogous to the videotape and audiotape treatments of the study reported here because sound recording of a classroom speech may well provide more information about the degree of demonstrated speech skill than the soundtrack of a training film provides about aerodynamics.

---

[1] Smode, "Learning and Performance."

[2] Bilodeau and Bilodeau, "Motor-Skills Learning," pp. 257-58.

[3] Harold E. Nelson, "The Relative Contribution to Learning of Video and Audio Elements in Films," Speech Monographs, XVIII (March, 1951), 70-73.

More analogous to videotape and audiotape replay is Kramar and Lewis' study in which auditors who could see a speaker comprehended significantly more than those who could only hear him.[1] Thompson provides a summary of quantitative research on differences in transmission of information by various media:

> Comparison of the media are of practical significance, and a certain lawfulness characterizes the findings. . . . in those [studies] dealing with comprehension, the superiority usually lies with a multichannel approach.[2]

Summarizing research on the effect of the complexity of stimuli transmitted to auditors, Thompson says that the weight of evidence "favors the value of presentations appealing to more than one of the senses."[3]

Taken together, the above findings suggest that the demonstration of skill is proportional to completeness of feedback which is influenced by the channel of communication and the complexity of stimuli it can transmit.

The previous discussion of the effect of feedback upon learning and performance applies equally to learning skill in speech performance. Karl Robinson describes one of the basic functions of the classroom teacher as that of providing performance feedback:

---

[1] Edward J. J. Kramar and Thomas R. Lewis, "Comparison of Visual and Nonvisual Listening," Journal of Communication, I (November, 1951), 16-20.

[2] Wayne N. Thompson, Quantitative Research in Public Address and Communication (New York: Random House, Inc., 1967), p. 161.

[3] *Ibid.*, p. 164.

> Each recitation offers a possibility for the improvement of the personal speech habits and speech skill of the student. The teacher must make judgments of such development through the various methods at his disposal. He must then reveal his findings to the student through a suitable type of criticism so that he can work on the things that need further development.[1]

But what is a "suitable type of criticism" for one student may not be so for another student. The relative uniqueness of students requires individualized treatment as Gilkinson discovered during his analysis of research on teaching speech skill: "The most successful teaching seems to be done in those classes in which the teacher makes a direct attack upon the specific problems of the individual student."[2] Since this quality of individualization of analysis is inherent in the private videotape or audiotape replay of a student's speech, these statements make it reasonable to suppose that this technique should be pedagogically effective.

One other unique characteristic of electronic feedback is that it permits the student, as no other technique can, to view an "objective" recreation of his own performance so that he can evaluate it himself. Ochs points up the importance of this when he says, "Of the three loci of classroom criticism--i.e. the instructor, a classmate, or the speaker--self-generated criticism is most effective in producing desired behaviors."[1] The importance of self-discovery and self-motivation has been stressed by Carl Rogers and his followers for years. Without accepting it as virtually the only means of changing behavior as Rogers might, many teachers of speech would agree that self-criticism is especially important in the process of learning speech skill.

---

[1] Karl F. Robinson, Teaching Speech in the Secondary School (New York: Longmans, Green & Co., 1951), p. 114.

[2] Howard Gilkinson, "Experimental and Statistical Research in General Speech: I. Effects of Training and Correlates of Speech Skill," The Quarterly Journal of Speech, XXX (February, 1944), 101.

The advantage of tape replay becomes clearer when one realizes that the alternative requires reliance upon the verbal descriptions, evaluations, and perhaps mimicking gestures and vocal inflections of a classmate or teacher. In such a situation, the student often fails to fully understand the nature of the criticism, or to fully accept its validity. With electronic replay, on the other hand, he can view the performance for himself. Even if his reaction to his own performance is not completely "objective," the precision and completeness of the feedback should lead to a more accurate picture and greater motivation than traditional criticism methods. This advantage can be seen in the statements of students who have had both traditional and tape feedback. As one said, "It's hard to be told what you are doing wrong, but it's easy to see it."[2] Besides realizing the information provided by videotape replay, students also seem to appreciate the motivational effect of evaluating themselves. As one put it, ". . . when criticized I'll agree, but when I see and hear my faults I'm more apt to do something about it."[1]

---

[1] Donovan J. Ochs, "Videotape in Teaching Advanced Public Speaking," Speech Teacher, XVII (March, 1968), 111.

[2] Chester Caton and George K. Feather, "Teaching Speech with Television," NAEB Journal, XXIV (November-December, 1965), 26.

In summary, feedback is essential to the acquisition and demonstration of a complex skill such as speech. Tape replay operates as performance feedback and therefore, it is hypothesized, provides an effective teaching aid in the area of speech. Further, there seems to be ample evidence for the theory that the greater the degree of completeness and accuracy of the feedback, the greater the degree of speech skill the student will demonstrate at the end of the training period. It is perhaps the implicit understanding of this theoretical model that has led teachers of speech to assume the pedagogical value of these replay techniques. However, this theory is of sufficient practical as well as theoretical importance to warrant its testing by means of experimental research. As Simon states:

> During its long life speech has accumulated diverse beliefs and assumptions, many of them from speculative or authoritarian sources. Efficiency in speech performance and in pedagogical practice demands the scientific testing of the tenability of these accumulated traditions.[2]

It is for this reason that this study was undertaken.

Organization of the Study

Chapter II surveys research related to two basic concerns of this study: (1) pedagogical uses of electronic feedback; (2) instruments for measuring speech skill. Chapter III describes two pilot studies conducted as part of this research project: (1) a test of the videotape recording and replay procedure; (2) a test of the instrument and procedure for measuring speech skill. Chapter IV treats the procedure employed in the basic experiment, including variables, controls, subjects, experimental activities, and procurement of data. Resulting data and statistical analysis is presented in Chapter V. Chapter VI draws conclusions from the analysis of experimental data and summarizes the findings of the study. Appendices reproduce all important sets of subject instructions and measurement instruments used and some of the resulting data.

---

[1] Nelson, "Instructional Uses of Videotape," p. 102.

[2] Clarence T. Simon, "Speech as a Science," Quarterly Journal of Speech, XXXVII (October, 1951), 292.

CHAPTER II

REVIEW OF RELATED RESEARCH

Research Dealing with Pedagogical Uses of Electronic Feedback

During the past thirty years an extensive number of articles dealing with the pedagogical utility of electronic feedback have appeared in speech publications. For the most part the authors have discussed the increase in student learning supposedly brought about by this method, without providing experimental evidence. The three means of electronic feedback employed in these studies are: sound recording, film, and most recently videotape.

Sound Recording

The earliest report of sound recording of student speech performances was given by Watkins in 1924.[1] Although he noted improvement in student grammar and delivery, he failed to indicate any attempt to quantify this apparent improvement or to compare the experimental group improvement to that of a control group. A decade later, Williamson described an extensive program using recording disks for student performance feedback.[1] The author reported enthusiastic acceptance of the technique by the majority of students and instructors involved. Similarly, Littell reported the value of recording student speeches to permit personal self-evaluation.[2] Neither Williamson nor Littell offered empirical evidence of the supposed improvement in student learning. Even the National Research Council's Committee on Scientific Aids offered no empirical backing for their endorsement of sound recording as a teaching aid.[3]

---

[1] Dwight E. Watkins, "An Apparatus for Recording Speeches," Quarterly Journal of Speech, X (1924), 253-58.

In what appears to be the earliest experimental study of the effect of sound feedback on learning speech skill, Nystrom and Leaf failed to disprove the null hypothesis.[4] Experimental subjects made a number of recordings of their speaking during the semester and listened to those recordings during weekly one-half hour periods without the presence of the instructor. Final readings of prose and poetry were rated by five competent judges on factors of pitch, rate, loudness, quality, enunciation, pronunciation, and general ability. These ratings showed no significant difference between the experimental group and the matched control group. This result may have been caused by the students' inability to perceive what aspects of their performance needed practice. Both Simons[1] and Robinson[2] recommend that a teacher using sound recording replays should make out a check list indicating what each student should listen for and practice.

---

[1] Arleigh Williamson, "Two Years' Experience with Recording Equipment," Quarterly Journal of Speech, XXI (October, 1935), 195-216.

[2] Winifred H. Littell, "Before and After Taking," Quarterly Journal of Speech, XXIII (December, 1937), 616-19.

[3] Harlen M. Adams, "Magnetic Tape Voice Recording," Quarterly Journal of Speech, XXVII (October, 1941), 379-82.

[4] Clarence L. Nystrom and Robert Leaf, "The Recording Machine as a Teaching Device," Quarterly Journal of Speech, XXV (October, 1939), 433-38.

Film

A number of studies have also been published on the use of silent and sound film for performance feedback. Like those on sound recording, these reports give enthusiastic support for the value of film as a method of teaching speech, without providing evidence other than positive student and instructor reactions. In 1937, Osband suggested that since film was widely used to correct the performance of football players, it might be appropriate as a corrective device in the field of speech.[3] Two years later, Utzinger reported on the use of a 16mm camera and simultaneous recording device to preserve portions of student speeches. The author attempted only general observations of the effectiveness of audio-visual playback, including the following:

> As the pictures are being shown [twice each], I make some comments, but they seem unnecessary, for all of the faults are so easily observed by the students themselves. . . . The student reaction to this experiment was splendid. Not only did it create a great amount of interest, but they all felt that it was

---

[1] I. F. Simmons, "The Nature and Use of Audio-Visual Aids in Speech Instruction," Southern Speech Journal, IX (November, 1943), 215-19.

[2] Karl F. Robinson, Teaching Speech in the Secondary School (New York: Longmans, Green & Co., 1951).

[3] Helen Osband, "The Giftie Gie Us," Southern Speech Journal, III (November, 1937), 19-22.

exceptionally helpful to them.[1]

Stanmyre published a report on film experiments at Syracuse University which praised the pedagogical use of sound film in the fundamentals of speech class.[2] Although original plans for the study included controlled quantification of results, this phase had to be dropped because of problems of equipment and student stage fright during preliminary testing. The three levels of feedback Stanmyre had hoped to test were sound film, disk recording, and no recording. During the basic study, students were filmed in a special studio having concealed equipment. Of the 870 students filmed during a three year period, 94% were favorable to the project as an aid to learning, although 82% expressed surprise at seeing themselves on the screen. All instructors agreed that the students benefited from the experience.

Eisenstadt described a similar study conducted at Rutgers University.[3] Here silent 8mm filming was done in class rooms. After processing, the films were shown to the entire class, once without interruption, and once with interruptions for comments by the instructor and students. Although the author reports greater motivation on the part of students as a result of this experience, no quantitative evidence is given for this conclusion.

In spite of its apparent value, the use of film has not gained wide use in the teaching of speech. Three reasons may be given for this: The first is the high cost of film which, unlike videotape and audiotape, can not be re-used. Second is the delay of several days between filming and projection caused by the need for processing the film. The final reason is the distraction caused by the noise of the camera and projector.

---

[1] Vernon A. Utzinger, "The Use of the Moving Picture Machine and the Recording Instrument in Teaching Speech," Quarterly Journal of Speech, XXV (February, 1939), 9-12.

[2] R. William Stanmyre, "The Syracuse Film Experiments," Quarterly Journal of Speech, XXXVI (February, 1950), 57-60.

[3] Arthur Eisenstadt, "As Others See Us," Speech Teacher, I (March, 1952), 121-24.

Videotape for Performance Feedback

As noted in Chapter I, videotape replay of student performances for self-evaluation has been accepted by many speech teachers. The studies which have dealt with videotape feedback fall into two categories: those which praise the technique as a means of increasing the degree of speech skill acquired in class without offering experimental evidence, and those experimental studies which have chosen to quantify something other than speech skill as the dependent variable. No published experimental evidence provides sufficient grounds for believing, as do so many teachers, that videotape self-confrontation increases the student's rate of acquisition of skill in speaking.

Among those writers praising videotape replay without giving appropriate evidence are Saltzman,[1] Henderson,[2]

---

[1] S. W. Saltzman, "TV Records Student Performances at University of Illinois," NAEB Journal, XXIII (March-April, 1964), 3-6.

[2] John Henderson, "Using Mirror TV to Teach Speaking," NAEB Journal, XXIII (November-December, 1964), 53-56.

Caton and Feather,[1] Nelson,[2] Ochs,[3] and Reynolds.[4] Ross[5] as well as Becker, Bowers, and Gronbeck[6] have experimented with the use of television "productional manipulative feedback" in teaching group discussion. In this approach, the instructor directs the production aspects of the videotaping. For example he may instruct a camera to zoom in on the mouth of a participant dominating the discussion, reveal the audience yawning, or show a discussant speaking in generalities with the word "support" superimposed below his picture. These tapes are then played back to the participants and class. No quantification has yet been published on the productional manipulative feedback technique.

The next group of researchers did bring quantification to bear upon the problem of the pedagogical value of television replay. One of the earliest of these attempts was conducted by Phelps and Hempen.[1] Rather than presenting the student's audio-visual image after the performance by means of videotape replay, they confronted the speaker with simultaneous transmission of the visual portion of his performance as it was given. This was accomplished by placing a television monitor in front of the speaker so he could view his performance from the camera's vantage point as he spoke. The authors report that this technique "did not lead to greater than normal progress in the mastery of the skills of physical delivery," since "speakers for good reason ignored the monitor and spoke instead to the audience."[2]

---

[1] Chester F. Caton and George K. Feather, "Teaching Speech with Television," NAEB Journal, XXIV (November-December, 1965), 24-26.

[2] Harold E. Nelson, "Videotaping the Speech Course," Speech Teacher, XVII (March, 1968), 101-03.

[3] Donovan J. Ochs, "Videotape in Teaching Advanced Public Speaking," Speech Teacher, XVII (March, 1968), 110-12.

[4] R. V. E. Reynolds, "Videotape in Teaching Speech in a Small College," Speech Teacher, XVII (March, 1968), 113-15.

[5] Raymond Ross, "The Wayne State University Program," The Student as Speaker and Listener, ed. by Elwood Murray and Arthur Solomon (Yellow Springs, Ohio: Antioch College and Jack Wolfram Foundation, 1966), 13-14.

[6] Samuel L. Becker, John Waite Bowers, and Bruce E. Gronbeck, "Videotape in Teaching Discussion," Speech Teacher, XVII (March, 1968), 104-06.

Another study of the impact of television feedback, this time in the form of performance replay, was conducted by Livingston and Doler.[3] According to Holladay, their findings provide evidence for the theory that videotape feedback has advantages over conventional oral criticism in teaching the basic course.[4]

One of the most widely quoted projects to investigate the use of videotape for self-evaluation by fundamentals of speech students was completed in 1966 by

---

[1] Phelps and Hempen, "Beginning Speeches."

[2] Ibid., p. 36.

[3] Harold M. Livingston and Thurston E. Doler, "Effectiveness of Video-Taped Feed-Back versus Conventional Oral Critique in Teaching the Basic College Speech Course" (unpublished paper, Oregon State University, 1964, mimeographed), as described by Howard Holladay, "A Pilot Study of the Effects of Simultaneous, Immediate, and Delayed Videotape Playback on Self-Perception of Delivery Skills" (unpublished paper, University of Southern California, Los Angeles, 1968), p. 2.

[4] Further information regarding this study is not available.

Adeline Hirschfeld.[1] Students from five sections of the basic speech course were videotaped twice during the semester, once during either the second or seventh week, and again during the last week. All speeches were rated by the instructor, speaker, and classmates during the immediate audiotape replay. Videotape replay was accomplished at another class meeting. Hirschfeld's primary concern was the effect of this technique upon speech ratings by the instructor, classmates, and the speakers themselves. These ratings seemed to show greater convergence when the videotaping activity was done during the second week than when it was done the seventh week. Unfortunately no statistical analysis of this data or quantification of the overall pedagogical value of electronic feedback was attempted. However the thoroughness of description provides insight into the process and problems of using videotape feedback.

Frandsen, Larson, and Knapp empirically tested the use of videotape replay as a "supporting device in the communication that flows from the instructor to the student regarding the student's recorded speaking behavior."[2] During the eighth and ninth weeks of a fundamentals of speech course, sixteen subjects were assigned to each of four treatment groups. Immediately after videotaping an assigned speech in the television studio, subjects received one of the following treatments: (1) "Before"--instructor's evaluative comments were given before the subject viewed the videotape replay of his speech; (2) "Simultaneous"--instructor's comments were given during the replay; (3) "After"--instructor's critique followed the replay; (4) "Control"--neither comments nor playback were given to the speaker. Present in the studio during this procedure were several television technicians, the instructor, and occasionally another student.

---

[1] Adeline G. Hirschfeld, "Utilization of Video Taped Speeches for Self Analysis in Fundamentals of Speech Course" (unpublished Ph.D. dissertation, Wayne State University, 1966).

[2] K. D. Frandsen, C. E. Larson, and M. L. Knapp, "Some Effects of Simultaneous and Sequential Video-Tape Playback and Instructor Comments: A Preliminary Report" (unpublished paper presented at the Conference of the Directors of the Basic Course of the Midwestern Universities, Lawrence, Kansas, February 18, 1967).

The dependent variable, correspondence between the instructor's judgment and the student's self-perception, was measured by having both instructor and student fill out a fifteen-item semantic differential immediately after the critique-playback treatment had been administered. This instrument contained four dimensions: competence, dynamism, success, and value. Although no discussion of the scale is presented by the authors, it appears to be similar to one developed by Smith.[1] Analysis of variance of mean differences between subject and instructor scores indicated that on two of the four dimensions, dynamism and success, no significant difference was found between the four treatment groups. On the value dimension, the "After" group showed significantly greater correspondence with instructor ratings than the "Control" group at the .05 level. On the competence dimension, both the "Simultaneous" and the "After" groups showed significantly greater instructor correspondence than either the "Before" group or the "Control" group at the .05 level. In no case was the group which received instructor critiques before viewing the videotape replay different from the control group. This study offers some evidence for the value of television feedback followed by instructor comments for increasing the degree of correspondence between the instructor's image of the student and the student's image of himself. However it provides no test of whether the technique increases student competence in speech skill.

---

[1] Raymond G. Smith, "Development of a Semantic Differential for Use with Speech Related Concepts," Speech Monographs, XXVI (November, 1959), 263-72; and "Validation of a Semantic Differential," Speech Monographs, XXVII (March, 1960), 50-55.

A recently completed pilot study by Holladay also investigated the effect of videotape feedback on the student's self image.[1] The independent variable was length of time between the performance and feedback. Three different sections of fundamentals of speech were assigned as units to the three treatments: (1) "simultaneous playback"--speakers were instructed to observe the video portion of their performance on a floor monitor while they spoke; (2) "immediate playback"--speakers were shown their performance replay the same class period it was recorded; (3) "delayed playback"--speakers were not permitted to see their performance replay until five days after it had been recorded. During this speech round, speakers received no written or oral criticism from either classmates or instructor. Immediately after viewing his playback, each speaker indicated his self-perception by filling out a rating form containing the dimensions of vocal delivery, visual delivery, and general effectiveness. Scores on these self-evaluations were compared to those of similar forms filled out by the speaker one week before his performance, and with ratings given by the instructor using the same instrument during the speech. It was assumed that if one of the three methods of playback were more effective in changing speaker self-image, that fact would be apparent in a comparison of group mean differences between evaluation scores. Speaker pretest and post-test self-evaluation scores were compared, as were differences between speaker post-test scores and ratings given by the instructor. Analysis of variance on both types of difference scores failed to indicate any significant difference between the three treatment groups. However the group mean differences were in the direction of suggesting that immediate playback resulted in greater shift of self-image and greater concurrence with instructor evaluations than the other two treatments.

---

[1] Howard Holladay, "A Pilot Study of the Effects of Simultaneous, Immediate, and Delayed Videotape Playback on Self-Perception of Delivery Skills" (unpublished paper, University of Southern California, Los Angeles, 1968).

Only preliminary results are available from an experimental study in speech pedagogy currently being carried out by McCroskey and Lashbrook.[1] The independent variable is feedback mechanism used in teaching fundamentals of speech. One complete class was assigned to each of three feedback conditions or treatments: (1) "Traditional"--students received criticism of their speeches from classmates and instructor in the normal manner; (2) "Video-tape recording only"--students viewed replays of their performances, but received no criticism from either classmates or instructor; (3) "Traditional + VTR"--students received feedback from class and instructor critiques as well as from videotape replays. One unusual feature of the videotape feedback was the use of a "special effects amplifier" which permitted the recording to show the speaker and the class-audience simultaneously on split-screen.

---

[1] James C. McCroskey and William B. Lashbrook, "Video Tape as a Self-Confrontation and Critical Device in the Teaching of Public Speaking" (incomplete, unpublished research described in phone interview, August 1, 1968).

Their experimental design calls for the dependent variable, "cognitive improvement" in speech, to be measured in several ways. One uses the average of judges' ratings of pretest and post-test self-evaluation essays in terms of their "degree of insight into the communication process." Another measure analyzes subject scores on pretest and post-test semantic differentials regarding the use of television for instruction, their speech course, and their own speaking. These first two parts of the study have not been completed at this writing.

The only measures tabulated and analyzed to date are speech grades and test scores. Their analysis demonstrated that the "Videotape recording only" group was significantly lower than the other two groups, at the .01 level. However, since one complete class made up each treatment group, the difference found may have been due solely to errors produced by the confounding of class with treatment.[1] Where the class is the unit for randomization rather than the individual student, a systematic difference between classes may be caused by such factors as meeting time, place, other students enrolled, and instructor to name only a few. Once classes begin operating as units, distinct "class personalities" often emerge having an immeasurable effect upon such measures of the dependent variable as speech grades and test scores.[1] While these problems of design and analysis do not mean that the findings of McCroskey and Lashbrook are incorrect, they do indicate the need for further research.

---

[1] E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education (Boston: Houghton Mifflin Company, 1958), p. 36.

A number of other experimental studies in feedback have been conducted by members of the field of speech, but none of these have direct bearing on the hypothesis of the research undertaken here.[2]

A large number of studies on videotape feedback, many of them experimental in nature, have been conducted by scholars in fields other than speech. Some offer secondary evidence for the hypothesis under investigation. The following examples are representative of their diversity, although not their number.

---

[1] Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally and Company, 1966), originally appearing as "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. by N. L. Gage, same publisher, 1963, pp. 171-248; William L. Hays, Statistics for Psychologists (New York: Holt, Rinehart, and Winston, Inc., 1963), pp. 471-89.

[2] Manuel M. Leonardo, "An Experimental Study of a 'Teaching Machine' Application to Speech Improvement" (unpublished Ph.D. dissertation, University of Michigan, 1963), as abstracted in Speech Monographs, XXXI (August, 1964), 235; Craig F. Johnson and Paul A. Games, "Effects on Learning of Adding 'Typical Questions' Responses to Television Lectures as a Type of Feedback," Speech Monographs, XXXII (November, 1963), 458-60; Thomas F. Moser, "An Experimental Study of the Effects of Verbal Reinforcement on Small Group Discussion" (unpublished Ph.D. dissertation, University of Michigan, 1965); Calvin N. Smith, "Evaluating Classroom Discussion" (unpublished paper presented at the Central States Speech Convention, Chicago, April 5, 1968); Donald Faules, "The Relation of Communicator Skill to the Ability to Elicit and Interpret Feedback under Four Conditions," Journal of Communication, XVII (December, 1967), 362-71.

Allen and Fortune's continuing research on "micro-teaching," a teacher training technique, has had far-reaching impact in education circles.[1] They describe this innovation:

> Micro-teaching is a scaled-down teaching encounter which has been developed at Stanford University to serve two purposes, (1) as preliminary experience and practice in teaching and (2) as a research vehicle to explore training effects under controlled conditions. In micro-teaching the trainees are exposed to the variables in classroom teaching without being overwhelmed by the complexity of the situation. They are required to teach brief lessons (5 to 10 minutes) in their teaching subject, to a small group of pupils (up to 5). These brief lessons allow opportunity for intense supervision, video-tape recording for immediate feedback and the collection and utilization of student feedback.[2]

An eight item "Teacher Demonstration Rating" form with five intervals per item was developed for evaluating trainee teaching performance. Four groups of teacher trainees were assigned to two treatments: Groups 1 and 2 received instruction by the micro-teaching method, groups 3 and 4 by the traditional instruction-theory method. Both pretest and post-test measures were taken for each group, permitting analysis of the dependent variable, teaching performance, by adjusted mean total post-test scores. Following an F ratio significant at the .01 level, Scheffe S-contrasts showed the micro-teaching groups significantly better than the traditional groups at the .01 level. A comparison of the pretest and post-test measures for the micro-teaching groups on each of the eight TDR factors indicated that during the training period they had made significant improvement on four: Understanding of aims, Method-communication, Reinforcement, and Global reaction. Method-communication showed the most significant improvement. No improvement was found on the other factors: Development of aims, Organization of content, Meaning of content, and Review. These findings suggest that micro-teaching has its greatest impact upon overt communication skills, its smallest upon cognitive skills. The authors summarize the overall findings saying simply, "Micro-teaching is a valid and effective means of preinternship training."[1]

---

[1] Dwight W. Allen and Jimmie C. Fortune, "An Analysis of Micro-Teaching: A New Procedure in Teacher Education" (unpublished paper, Stanford University, February 8, 1965, mimeographed).

[2] Ibid., p. 1.

Spitzer, Jackson, and Satir conducted research on the use of videotape replay for training psychiatrists, psychologists, and psychiatric social workers in conjoint family therapy.[2] Trainees viewed videotapes of their family therapy sessions, much as Allen's interns viewed their micro-teaching sessions. Although the authors give no quantification, they report observing improvement in therapy techniques as a result of this teaching approach:

> This technique makes possible a heightened emphasis on nonverbal communication, particularly that

---

[1] *Ibid*., p. 10.

[2] Robert S. Spitzer, Don D. Jackson, and Virginia Satir, "A New Technique for Teaching Conjoint Family Therapy," *Psychiatric Spectator*, I (1964), 13-14.

> which goes on outside of awareness. The trainee-therapist sees himself on television, which encourages him to learn to see himself as others see him, as well as to explore the total effect he is having on others. . . . It is suggested that the use of this particular tool can be helpful in making the data of psychotherapy and research both more interesting and objective.[1]

The therapeutic effect of videotape feedback was investigated by Moore, Chernell, and West.[2] In their research, eighty mentally ill patients were videotaped during their interview of admission to the University of Mississippi Medical Center. One half of these patients viewed replays of their own interviews. At the end of the period of treatment a substantially greater number of patients from the videotape feedback group were judged "maximally improved clinically" than from the control group.

Another example of research outside the area of speech is provided by Haines in his report of videotape simulation and self-confrontation in training USAF Commandos for culture-contact and interaction skill.[3] The procedure involves six steps: (1) the trainee is given a short lecture on how to work with his counterpart in the host country; (2) video tape is shown depicting successful use of cross-cultural skills; (3) another tape is shown depicting a steadily worsening cross-cultural scene; (4) at the critical moment, the tape is stopped and the trainee is asked to role-play the completion of the situation to resolve the problem, while his performance is recorded on videotape; (5) the trainee is immediately confronted with the playback of his role-playing behavior; (6) during the playback, a skilled instructor critiques the performance. Although the author reports the success of this training procedure, no empirical evidence is provided.

---

[1] *Ibid.*

[2] Floyd Jack Moore, Eugene Chernell, and Maxwell J. West, "'O Wad Some Power the Giftie Gie Us': Television as a Therapeutic Tool," *Psychiatric Spectator*, I (November, 1964), 14-15.

[3] Donald B. Haines, "Training for Culture-Contact and Interaction Skills," AMRL-TR-64-109 (Wright-Patterson Air Force Base: Behavioral Sciences Laboratory, Air Force Systems Command, 1964).

In summary, a number of studies have been reported on the pedagogical value of electronic feedback. While several writers have praised the use of sound recordings for speech student self-evaluation, the single experimental study dealing with this method of feedback showed no difference between the experimental class and the control class. Several studies have recommended the use of motion picture film to preserve student performances, but none have offered empirical evidence supporting their assertions.

In the brief time videotape has been available, a number of writers have advocated it as an adjunct to traditional methods of teaching speech. However most of these reports have been based on personal observations rather than controlled comparisons and reliable measurements. The few studies which have employed experimental methodology have failed to investigate the question of whether videotape feedback increases the student's rate of acquiring skill in speaking. The quality of investigation by educational researchers into the use of videotape feedback for teacher training points up the need for speech scholars to conduct similar quantitative research on its use for speaker training.

Research Dealing with Instruments for Measuring Speech Skill

With the growth of experimentation on teaching speech and on composing and presenting speeches has come a need for reliable, valid means of measuring the effectiveness of speech performance. While there are many reasons for the absence of a universally accepted method of measurement, one appears to be the view that measuring speech performance is so impossibly complicated that no device can be both objective and valid. However, modern analytical techniques of factor analysis, advances in psychometrics, and more extensive research using and testing measurement instruments have led to increased confidence in the feasibility of measuring speech skill effectively. The remaining sections of this chapter will discuss specific research which has led to this increased confidence, and the best available instrument for measuring the dependent variable in this study.

A wide variety of measurement devices has been employed in speech research.[1] While these diverse methods may be classified in many ways, it seems meaningful to discuss them here in terms of two general categories: Those which seek to quantify the auditor's reaction to a speech in an indirect or projective manner, and those which ask the auditor for a direct evaluation of the speech. The more indirect methods have been less widely employed in speech research, and will be discussed first.

---

[1] Howard Gilkinson, "Experimental and Statistical Research in General Speech," Quarterly Journal of Speech, XXX (February, 1944), 95-101 (and April, 1944), 180-86.

Indirect or Projective Instruments

In a recent study, Cronkhite investigated the levels of autonomic arousal, heart rate, and skin conductance, while his subjects listened to a persuasive speech and completed an attitude test relating to the speech topic.[1] He reports that "subjects who changed attitudes most in response to the speech did not show greater GSR frequency increase at any time," nor did they show significantly greater heart rate.[2] However, there is no way to determine whether these unvalidated measures of speech persuasiveness, the dependent variable, are inadequate, or whether the research hypothesis is simply untenable. A similar although less clinical approach was taken by Kretsinger when he used gross bodily movement as an index of audience interest in a theatrical production.[3]

Objective recall tests based on material presented in speeches have been used to investigate relative effectiveness of different presentations. Frandsen used this procedure,[4] as did Ehrensberger,[5] and Gulley and Berlo.[6]

---

[1] Gary L. Cronkhite, "Toward a Real Test of Dissonance Theory," Quarterly Journal of Speech, LII (April, 1966), 172-78.

[2] *Ibid*., p. 178.

[3] Elwood A. Kretsinger, "An Experimental Study of Gross Bodily Movement as an Index to Audience Interest," Speech Monographs, XIX (November, 1952), 244-48.

[4] Kenneth D. Frandsen, "Effects of Threat Appeals and Media of Transmission," Speech Monographs, XXX (June, 1963), 99-107.

[5] Ray Ehrensberger, "An Experimental Study of the Relative Effectiveness of Certain Forms of Emphasis in Public Speaking," Speech Monographs, XII (1945), 94-111.

[6] Halbert E. Gulley and David K. Berlo, "Effect of

Nelson employed the recall procedure to test relative effectiveness of different channels of communication.[1]

The most promising projective measure of speech skill is the class of instruments known as semantic differentials. These instruments are designed to quantify the connotative meanings of words by means of a series of bipolar adjective scales.[2] Theoretically the connotative meaning of a given concept may be represented by a point in a multi-dimensional "semantic space." Early work on a general semantic differential established three basic dimensions: evaluative, potency, and activity. However, these factors may not be considered applicable for the measurement of any particular concept by any particular judges. Osgood points out that his procedure is "a highly generalizable technique of measurement which must be adapted to the requirement of each research problem to which it is applied."[3] Factor analysis offers the most widely accepted means of devising an instrument for a specific measurement problem.

Tannenbaum employed a semantic differential to study the impact of music on auditors' judgments of stage and televised drama.[4] Andersen successfully used a specially

---

Intercellular and Intracellular Speech Structure on Attitude Change and Learning," Speech Monographs, XXIII (August, 1956), 288-97.

[1] Harold E. Nelson, "The Relative Contribution to Learning of Video and Audio Elements in Films," Speech Monographs, XVIII (March, 1951), 70-73.

[2] Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Illinois: University of Illinois Press, 1957).

[3] Ibid., p. 76.

[4] Percy H. Tannenbaum, "The Effect of Background Music on Interpretation of Stage and Television Drama," Audio-Visual Communications Review, IV (Spring, 1956), 92-101.

constructed semantic differential based on factor analysis to quantify the dimensions of several prominent personalities' ethos.[1] Gulley and Berlo attempted to measure audience shift of attitude by means of several semantic differential evaluative scales.[2] Evaluative scales also made up the instrument used by Bettinghaus to "specify the interaction of five variables believed responsible for attitude change in a public speaking situation: (1) the speaker, (2) the speaker's delivery, (3) the speech topic, (4) the treatment of the speech topic, and (5) the assertion linking speaker and speech topic."[3] Based on results from analysis of data, he was able to state: "In terms of gross effects, 'effective' delivery in this experiment produced a more favorable shift in attitude toward the speaker than did 'ineffective' delivery."[4]

Smith has attempted to apply Osgood's technique of scale development "to speech concepts and to subjects who had been subjected to varying amounts of speech training in an attempt first to identify and define the dimensions along which speech concepts are judged, and as a consequent [sic] to make available a new and useful measuring instrument for speech critics and experimental research students."[5]

---

[1] Kenneth E. Andersen, "An Experimental Study of the Interaction of Artistic and Non-artistic Ethos in Persuasion" (unpublished Ph.D. dissertation, University of Wisconsin, 1961).

[2] Gulley and Berlo, "Effect of Structure."

[3] Erwin P. Bettinghaus, "The Operation of Congruity in an Oral Communication Situation," Speech Monographs, XXVIII (August, 1961), 131.

[4] *Ibid*., p. 142.

[5] Raymond J. Smith, "Development of a Semantic Differential," 263.

Although he attempted to validate the resulting instrument for measuring speech skill, he failed to factor analyze data for this specific measurement situation.[1] This omission is especially damaging in light of Clevenger's empirical findings that Smith's semantic differential for measuring reactions of a variety of audiences to different types of plays fails to achieve the broad usefulness claimed.[2]

Unfortunately the one semantic differential specifically designed to measure speech skill has not undergone factor analysis.[3] Although the semantic differential offers great promise in the measurement of speech performance, development of a valid instrument for that specific purpose has not been completed.

Direct or Evaluative Instruments

Instruments in this category require the auditor to make a direct measure of his reaction toward a speech in terms of some criteria. One common device of this type is the Woodward shift of opinion ballot,[1] exemplified in the research of Luchins,[2] and Frandsen.[3] However some writers have argued that apparent shift in audience attitude does not provide an accurate measure of speech skill.[4]

---

[1] Raymond G. Smith, "Validation of a Semantic-Differential," 50-55.

[2] Theodore Clevenger Jr., Margaret Leitner Clark, and Gilbert N. Lazier, "Stability of Factor Structure in Smith's Semantic Differential for Theatre Concepts," Quarterly Journal of Speech, LIII (October, 1967), 241-47.

[3] John Waite Bowers (unpublished scale, developed at the State University of Iowa, ca. 1964; discussed in a telephone interview, October 20, 1967).

Most direct or evaluative devices fall under the heading of rating scales. As Guilford points out, "Of the psychological-measurement methods that depend upon human judgment, rating-scale procedures exceed them all for popularity and use."[5] Remmers defines rating in the words of Good: "an estimate, made according to some systematized procedure, of the degree to which an individual person or thing possesses any given characteristic. . . ."[6] Becker discusses the speech rating phenomena:

> The rating of communicative behaviors (such as discussion) or of the artifacts of those behaviors (such as essays) is generally thought to be useful only for pedagogical purposes. This is unfortunate because, properly used, rating can be an important tool for communication research. By rating, I mean describing an individual, group of individuals, behaviors, or

---

[1] Howard S. Woodward, "Measurement and Analysis of Audience Opinion," Quarterly Journal of Speech, XIV (February, 1928), 94-111.

[2] Abraham S. Luchins, "Primary-Recenty in Impression Formation," Order of Presentation in Persuasion, ed. by Carl I. Hovland (New Haven: Yale University Press, 1957), 33-61.

[3] Frandsen, "Effects of Threat Appeals."

[4] Gilkinson, "Experimental and Statistical Research" (April, 1944), 186.

[5] Joy P. Guilford, Psychometric Methods (2nd ed. rev.; New York: McGraw-Hill Book Co., Inc., 1954), p. 263.

[6] H. H. Remmers, "Rating Methods in Research on Teaching," Handbook on Research on Teaching, ed. by Nathaniel L. Gage (Chicago: University of Chicago Press, 1967), Ch. 7.

> artifacts of behaviors in terms of one or more numerical scales. . . . Thus, through rating, one can obtain measures of such varied concepts as leadership, voice quality, degree to which the style of a message is adapted to the interests or needs of its intended audience, speech fright, the cooperation within a group, or the balance or attractiveness of a visual message.[1]

Many studies employing rating scales to quantify speech skill have used the single dimension, "general effectiveness." Among these are studies conducted by Borchers,[2] Gilkinson,[3] Clevenger,[4] and Frandsen.[5] Ordinal scaling, or ranking, on general effectiveness has been recommended by Fotheringham,[6] and Sawyer,[7] while the method of rank-comparison has been proposed by Bittner and Rundquist.[8]

---

[1] Samuel L. Becker, "Rating Scales" (unpublished chapter draft, dittoed, 1967), p. 1.

[2] Gladys L. Borchers, "Speech Without Work?" Quarterly Journal of Speech, XXI (June, 1935), 376-78.

[3] Howard Gilkinson and Franklin H. Knower, "A Study of Standardized Personality Tests and Skill in Speech," Journal of Educational Psychology, XXXII (March, 1941), 161-75.

[4] Theodore Clevenger, Jr., "Retest Reliability of Judgments of General Effectiveness in Public Speaking," Western Speech, XXVI (Fall, 1962), 216-22; and "Influence of Scale Complexity on the Reliability of Ratings of General Effectiveness in Public Speaking," Speech Monographs, XXXI (June, 1964), 153-56.

[5] Kenneth D. Frandsen, "Effects of Threat Appeals and Media of Transmission," Speech Monographs, XXX (June, 1963), 99-107.

[6] Wallace C. Fotheringham, "A Technique for Measuring Speech Effectiveness in Public Speaking Classes," Speech Monographs, XXIII (March, 1956), 31-37.

[7] Thomas Sawyer Jr., "A Grading System for Speech Classes," Speech Teacher, IX (January, 1960), 12-15.

[8] Reign H. Bittner and Edward A. Rundquist, "The Rank-Comparison Rating Method," Journal of Applied Psychology, XXXIV (1950), 171-77.

The majority of rating scales used in speech research ask the observer to evaluate the performance in terms of more than one characteristic or factor. A survey of literature in the field indicates that as few as three factors,[1] and as many as twenty-eight factors have been employed in various rating scales.[2] The largest number of orthogonal dimensions established by factor analysis is six.[3] In terms of the number of discriminations or intervals on each factor, reported research has shown a range of from two points on the Monroe, Remmers, and Venemann-Lyle check-list,[4] to twenty-one points on Bryan and Wilke's scales.[5] Research by Becker and Cronkite suggests that the scale length of five points provides measurements of highest reliability when the raters are beginning speech students.[6] However no research has dealt with the optimum scale length for highly motivated, experienced faculty judges.

---

[1] Samuel L. Becker, "The Rating of Speeches: Scale Independence," Speech Monographs, XXIX (March, 1962), 38-44.

[2] Wilbur E. Moore, "Factors Related to Achievement and Improvement in Public Speaking," Quarterly Journal of Speech, XXIX (April, 1943), 213-17.

[3] William K. Price, "The University of Wisconsin Speech Attainment Test" (unpublished Ph.D. dissertation, University of Wisconsin, 1964).

[4] Alan H. Monroe, H. H. Remmers, and Elizabeth Venemann-Lyle, "Measuring the Effectiveness of Public Speech in a Beginning Course," Studies in Higher Education, XXIX (September, 1936), 5-29.

[5] Alice I. Bryan and Walter H. Wilke, "A Technique for Rating Public Speeches," Journal of Consulting Psychology, V (March, 1941), 80-91.

[6] Samuel L. Becker and Gary L. Cronkite, "Reliability as a Function of Utilized Scale Steps," Speech Teacher, XIV (November, 1965), 291-93.

The evolution of the multi-factor speech rating scale may be traced through the following studies which have sought to develop and/or evaluate rating scales: Stevens,[1] Knower,[2] Norvelle,[3] Monroe, *et al*.,[4] Bryan and Wilke,[5] Thompson,[6] Fotheringham,[7] Brooks,[8] Becker,[9] and Price.[10] Of these only Becker and Price have used factor analysis to determine the independent dimensions represented by various scale items. Becker found that his eleven scales collapsed into three orthogonal dimensions: analysis-content, delivery, and language. Although he recommends that an instrument be based on this factor analysis, none has been built and tested for reliability or factor weights. On the other hand, Price has tested the instrument based on his own factor analysis, finding reliability estimates and weights for each factor. A factor analysis of responses on thirty-five scale items led to the establishment of six factors: speech content, intelligibility, bodily action, personality, language, and voice. Because of its unique completeness, the Price speech performance rating scale was selected to quantify the dependent variable in this study.

---

[1] Wilmer E. Stevens, "A Rating Scale for Public Speakers," *Quarterly Journal of Speech*, XIV (April, 1928), 223-32.

[2] Franklin H. Knower, "What is a Speech Test?" *Quarterly Journal of Speech*, XXX (February, 1944), 485-93.

[3] Lee Norvelle, "Development and Application of a Method for Measuring the Effectiveness of Instruction in a Basic Speech Course," *Speech Monographs*, I (1934), 41-65.

[4] Monroe, *et al*., "Measuring the Effectiveness."

[5] Bryan and Wilke, "Technique for Rating."

[6] Wayne Thompson, "Is There a Yardstick for Measuring Speaking Skill?" *Quarterly Journal of Speech*, XXIX (February, 1943), 87-91.

[7] Fotheringham, "Technique for Measuring."

[8] Keith Brooks, "The Construction and Testing of a Forced Choice Scale for Measuring Speaking Achievement," *Speech Monographs*, XXIV (March, 1957), 65-73.

[9] Becker, "Rating of Speeches."

[10] Price, "Speech Attainment Test."

The Price Speech Performance Rating Scale

The origination of the University of Wisconsin's honors program in 1959 led to a decision by the speech department to develop a testing procedure to recognize superior students registered for the basic speech course so they might be enrolled in a more advanced course. For that purpose, a two part "Speech Attainment Test" was constructed. "Part I--Concepts and Principles" was a paper and pencil test comprised of 150 multiple-choice items based on subject matter of the beginning speech course. All students who registered for basic speech were to take this test. Only those whose score was above a specific cutoff score would take the "Part 2--Oral Performance." Here they were asked to prepare a three minute persuasive speech which was rated by a group of faculty members and graduate assistants of the speech department.

In order to construct his scale, Price began with a list of thirty-five terms taken from the literature on speech rating scales. These were listed without definitions in random order on the experimental instrument, each accompanied by a 7-point scale. For the purpose of factor analysis of these thirty-five terms, an experiment was conducted in which kinescoped performances of twenty-six undergraduate students enrolled in an advanced course in persuasion were evaluated. The judges were eleven faculty members and twenty-three teaching assistants of the speech department. This process of rating was accomplished in a three hour period which included one fifteen minute rest period. The data collected were cast into a two-dimensional scheme with the axes being the speakers and the items of the rating scale. These observations were subjected to a "component analysis" leading to six factors, which in turn were subjected to Varimax rotation. The six factor designations assigned by Price, along with their major sub-factor components, are: (1) Content--reasoning, supporting material, concreteness; (2) Intelligibility--articulation, pronunciation; (3) Bodily action--movement, gesture, facial expression, eye contact; (4) Personality--personal appearance, poise, fluency; (5) Language--figurative language, word choice, originality of ideas; (6) Voice--vocal expressiveness, pitch, vocal quality, rate. The final test instrument is shown, with one modification, in Appendix A. Scale lengths have been expanded from seven steps per factor to eleven steps, in accordance with findings of the measurement procedure pilot study described in Chapter III.

For the purpose of selecting the University of Wisconsin students whose performance received a sufficiently high rating on this scale to enroll in an advanced speech course, a single score was needed as an over-all measure of each student's speech performance ability. Although standard practice has been to sum across all factor scores, Price correctly points out that this method is based on the false assumption that each factor is of equal importance in determining speech skill. Heinberg has experimentally demonstrated the invalidity of this assumption. He summarizes, "Hence, delivery is far more influential than is content in determining the general effectiveness of these two types of speeches."[1] Similarly, the Bettinghaus study mentioned earlier provides evidence supporting the greater impact of delivery on listener attitude.[2]

Price used multiple regression analysis to determine different weights for the six factors. Data for this analysis were gathered during the final week of two fundamentals of speech classes. Ten teaching assistants from the University of Wisconsin Department of Speech used the Price scale to rate twenty-seven speakers giving three minute persuasive speeches. The criterion measure adopted for the analysis was the numerical sum of the grades these students received from their instructor on the five speaking assignments given in that basic speech course. The multiple correlation coefficient between the six factors of the scale and the criterion measure was .97, indicating that 95% of the variance of the criterion measure is accounted for by the phenomenon measured by the rating scale.

Reliability of the scaling procedure was determined by evaluating six speeches of students who had received

---

[1] Paul Heinberg, "Relationships of Content and Delivery to General Effectiveness," Speech Monographs, XXX (June, 1963), 105-07.

[2] Bettinghaus, "Operation of Congruity."

γ

high scores on the written "Concepts and Principles" test and six who had received low scores. Reliability coefficients were computed for ratings of the 1-k judges employed, using an analysis of variance technique. Indicative of the generally high reliabilities found was the mean coefficient of .77 for three faculty judges on all six factors. Validity was also considered, but no attempt was made to compute a validity coefficient for measurements using the scale.

The value of the Price scale for the study reported here lies in its completeness of testing and the fact that it was designed and developed using beginning university speech students and speech faculty members. In this study, performances by a similar group of subjects were evaluated by a similar group of judges.

Summary of Research

Electronic Feedback

The earliest speech research on the pedagogical utility of electronic feedback employed sound recording devices to replay students' performances. In nearly all cases, the researcher describes his recording-playback procedure and gives his unquantified observations of the resulting increase in student learning. The single experimental study found no difference between final presentations of the experimental class which had heard recordings of their speeches and the control class.

Several studies have praised film as a means of replaying student speeches for self evaluation. However none provided experimental testing of the pedagogical value of film. It appears that the high cost, the disturbing noise, and the time delay between filming and screening have kept this technique from being adopted in more fundamentals of speech programs.

A large number of studies have been reported dealing with the application of videotape replay to teaching speech. Like previous research on electronic feedback, the majority of these have claimed that the technique has real pedagogical value, but have failed to provide experimental evidence. Several studies have attempted to quantify some aspect of the technique such as resulting congruence with instructor critiques or cognitive improvement in speech. However none has attempted to test the fundamental question, "Does videotape feedback help students achieve greater skill in speaking than they might acquire in a speech class not employing this technique?"

Outside the area of speech, research has been conducted on the use of videotape feedback to teach such complex skills as conjoint family therapy and cross-cultural communication skills. It has also been studied as a therapeutic aid for mental patients. The teacher training procedure known as "micro-teaching" has been shown experimentally to be more effective than traditional instruction-theory methods. The clarity of these findings is in marked contrast to those of speech studies dealing with videotape.

Measurement Instruments

A wide variety of devices has been employed in measuring aspects of speech skill for research purposes. These may be grouped in two categories: indirect or projective methods, and direct or evaluative methods.

Indirect methods have measured levels of autonomic arousal, gross bodily movement, shift of opinion, communication recall, and connotative meaning. This last index is quantified by the most promising of the indirect measurement devices, the semantic differential. However no factor analyzed, tested semantic differential is presently available for the specific task of measuring speech skill.

Shift-of-opinion ballots and rating scales of different types make up the second category, direct or evaluative methods. In speech research these are more traditional and more widely used than indirect devices. Ordinal rating of general effectiveness includes simple ranking and rank-comparison methods. Among interval rating scales are uni-factoral or "general effectiveness" scales and multi-factoral scales containing as many as twenty-eight apparent dimensions. Multi-factoral scales of three to eight factors have been most popular. In order for a rating scale to be acceptable for a particular measurement problem, it should be based on factor analysis to insure independence of factors. It should also have been tested on a population similar to those subjects to be measured. Because it best meets these criteria, the Price scale was chosen to quantify the dependent variable of this study.

CHAPTER III

PRELIMINARY TESTING OF PROCEDURES AND APPARATUS

Pilot Test of Videotape Replay Procedure

Problem and Testing Procedure

In the spring of 1966, a pilot study was undertaken to investigate several questions dealing with the process of videotaping and replaying student performances for self-evaluation. Answers to four questions were sought in this test of the videotape feedback procedure: (1) Will the procedure be feasible with the television recording facilities and staff available? (2) How valuable will students regard videotape feedback to be? (3) Will it take more student or instructor time than an activity of similar apparent value? (4) What improvements should be made in the taping-replay-evaluation procedure to increase its ease of operation or value?

Subjects for the pilot study were fourteen members of one Speech 124 class, the three-hour basic speech course at Eastern Michigan University. This investigation took place during the seventh or next-to-last round of speaking, while the class was in the Quirk TV studio for six days of closed circuit television speeches. Subjects were not informed that their speeches would be videotaped or that they were participants in a research project. However class members filling production jobs soon realized that speeches were being videotaped.

After giving his speech, each subject received written instructions to view his performance in Roosevelt TV studio at one of two specified times. Once there he viewed a classmate's speech once, then his own twice. Neither his instructor nor other students were present during these playbacks. His instructions were to take careful notes on his own performance following a critique form provided. Using these notes, he was required to write a self-evaluation and hand it in at the next class meeting. The critique form provided was essentially the same as the one printed on the back of student outline forms used by the instructor throughout the semester for evaluating and commenting on student performances.[1] It contained two major factors, "Delivery" and "Composition." The former was subdivided into directness, vitality, voice, and bodily action; the latter, into topic and speech purpose, organization, introduction and conclusion, supporting material, and language. The instructor's evaluation and grade were made available to each student only after he had handed in a self-evaluation.

Two means were used to provide answers to the four questions noted earlier. First, a 26-item questionnaire was administered to detect the reactions of the subjects toward the value of the procedure and ways in which it might be improved.[2] To increase candor in responses, subjects were asked to hand in questionnaires unsigned to a member of the class who checked off their names so they could receive credit. Fourteen subjects submitted questionnaires indicating they had done all pilot project activities. Second, several conferences were held between the investigator and the Coordinator of Instructional Broadcasting to discuss problems, possible improvements, and ways of implementing the procedure in the major experiment.

---

[1] See Appendix B for the complete critique form.

[2] See Appendix C for the complete text and responses.

Results

The first question regarding the feasibility of recording and replaying speeches with the available staff and facilities was answered affirmatively. The Coordinator of Instructional Broadcasting indicated that the use of Roosevelt TV studio for playbacks had proven no burden in terms of staff or equipment. Also, a total of 78% of the subjects indicated that the basic playback facilities in the studio were satisfactory by responding negatively to the question, "Did viewing the tape in Roosevelt #108 prove distracting?" Seventy per cent said they had found the engineers "helpful." However there was evidence that greater care would have to be taken in the training of student technicians who handled recordings and playbacks of the speeches. One performance had been recorded without audio, reducing the feedback to silent video replay. Also, 42% of the subjects indicated they had not seen their own performance twice and that of another student once as had been planned. Training personnel more carefully for these tasks, freeing them from other duties, and providing a careful system of checks on procedure seemed to offer the best solution to these problems.

No attempt was made to measure the actual effect of videotape feedback on student self-awareness or speech skill. However student responses on self-evaluations and questionnaires showed they felt videotape replay was highly valuable, providing an affirmative answer to the second question posed. Ninety-two per cent responded either "Definitely Yes" or "Yes" to the question, "Did this experience make you more aware of your problems in oral communication?" A total of 80% felt the videotape feedback procedure used was "valuable for delivery," while 56% said it was "valuable for composition." Seventy-seven per cent said they would have gone to see their videotape even if viewing had not been required. Finally, 100% indicated they thought that "video taping is generally a good technique in a fundamentals of speech course." As one student said in his self-evaluation, "I feel that this video tape was very valuable to me. It's hard to be aware of your problems until you can see them yourself."

The third question dealt with whether videotape replay would take more student or instructor time than an activity of similar apparent value. The evidence available provided a negative answer. In spite of the fact that 61% of the subjects said they had spent over one hour viewing the replay and writing a self-evaluation, 100% said that this use of videotape was worth their time spent in the activity. Although the instructor did not record his time spent in this activity, it did not seem to have been more than for an activity of similar value. Of course the technique did require extra time of Instructional Broadcasting personnel.

Question four dealt with improvements which might be made in the recording-replay technique. On the basis of the available data, major aspects of the method were left unchanged. The replay of a classmate's recording was used to "validate" the accuracy of the audio-video playback in order to increase the subject's acceptance of his own replay as a faithful representation of how he looked and sounded. This purpose seemed to have been fulfilled; 92% of the subjects responded affirmatively to the question, "Do you think that the TV picture and sound gave a fair representation of how you looked and sounded?" Also, 57% felt it was "helpful to see the performance of another classmate." Since research had suggested that a subject's defense mechanisms might blunt the effect of a single replay of his speech, it was decided to show each speaker his performance twice.[1] Although students were not asked whether viewing their speech once would have been preferable, their responses indicated that seeing it twice seemed appropriate. Fifty-four per cent responded negatively to the question, "Do you think it would be helpful to see your performance more than twice?" while 38% answered affirmatively. Written self-evaluations had been made a part of the procedure because past experience had suggested they would increase the student's awareness and internalization of the information and reinforcement provided by the feedback. In spite of the time and effort required in writing self-evaluations, 77% accepted the value of this activity, responding affirmatively to the question, "Did writing a self-evaluation make the video tape experience more meaningful?" This was the first assignment in which written self-critiques had been required.

---

[1] Adeline Hirschfeld, "Utilization of Video Taped Speeches for Self Analysis in Fundamentals of Speech Course" (unpublished Ph.D. dissertation, Wayne State University, 1966).

Students were asked, "If this speech recording had been done in the classroom do you think that the presence of the camera would have been distracting?" Their responses were evenly divided with roughly one-third affirmative, one-third undecided, and one-third negative, suggesting some anxiety over the thought of television equipment and personnel operating openly in the classroom.

On one matter, the feelings of subjects were not followed in the design of the master experiment. In spite of the fact that 79% felt it would have been "helpful to have the instructor's written critique" of their speeches while they viewed the replays, it was decided not to make those critiques available until after the subjects had handed in their self-evaluations. This decision took into account two important considerations: First, findings of Frandsen, Larson, and Knapp indicated that the greatest concurrence between student self-evaluations and instructor evaluations occurred when instructor critiques were given after students had viewed their videotaped performance.[1] Their test situation did not delay the instructor critiques by a day or more after replay, as was planned for this study. Therefore these results could only be taken as suggestions that subjects need not receive critiques before viewing their videotape in order for the replay to be helpful. Second, the fact that students could not get their instructor's comments and grade until they had turned in self-evaluations was used as an inducement to complete the experimental activities of viewing and evaluating.

---

[1] K. D. Frandsen, C. E. Larson, and M. L. Knapp, "Some Effects of Simultaneous and Sequential Video-Tape Playback and Instructor Comments: A Preliminary Report" (unpublished paper presented at the Conference of the Directors of the Basic Course of the Midwestern Universities, Lawrence, Kansas, February 18, 1967).

When all eight sections of Speech 124 were asked if they would have viewed videotape recordings of their first and third speeches had they been available, 85% said they would have, even if they had not been required to. Fully 92% of the students in those eight sections felt that "it would have been helpful" to have some of their speeches videotaped so they "could see how . . . [they] looked and sounded."

In summary, this pilot study demonstrated that it was feasible to videotape and replay student speeches using the procedure planned, and that students responded so enthusiastically that the extra time required seemed well spent.

Pilot Test of the Instrument and Procedure for Measuring Speech Skill

Problem and Testing Procedure

A pilot study was undertaken in the spring of 1968 to answer three specific questions regarding the Price scale and the procedure for employing it: (1) Will the training period and the judging session be practical and efficient? (2) Will the precision of measurement with the Price speech performance rating scale be increased by expanding scale lengths on individual factors from seven intervals per item to eleven intervals? (3) Will the combined number of student and faculty judges planned for the major experiment give ratings of sufficiently high reliability? None of the judges used in this pilot study were scheduled for the major experiment, although two were ultimately used in both.

The basic criterion for selecting faculty and student judges for the pilot study was experience in teaching or taking courses in public address. All six faculty judges had taught fundamentals of speech courses, four at Eastern Michigan University and two at Ann Arbor High School. All twenty student judges had had at least three courses in public address and group communication at Eastern Michigan University. In order to obtain this number of subject-judges, the pilot study was run on two different nights. It took place in Quirk #94, a room with no windows, equipped for audio-visual activities, with special lighting, projection screen, and television monitor. During the two replications, an attempt was made to hold all aspects of procedure constant. To eliminate any uncontrolled variables operating on the replications, a simple Latin square design was used.[1]

The procedure consisted of two separate phases, a training phase to help the judges learn how to use the Price scale effectively, and a judging phase to test their ability to use it. The training phase itself consisted of two parts: The first part began with a statement by the investigator that the reason for the project was, "To see how precisely we can measure 'speech skill' using the

---

[1] E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education (Boston: Houghton Mifflin Company, 1958).

University of Wisconsin Speech Performance Rating Scale." This was followed by a brief lecture with projected graphs illustrating three common errors of judging--central tendency, halo effect, and leniency effect.

The second part involved judging and discussing each of four videotaped student speeches. The four speeches had been selected by the investigator as examples of the best, worst, and average quality to be evaluated in the judging session, but this fact was not communicated to the judges. Speeches were viewed on a 23" RCA #23EV classroom TV receiver with sound channeled through an Ampex #620 high fidelity amplifier-speaker. This system was positioned at the front of the room, where the speaker lectern is normally located. For viewing all room lights were turned off except for low-intensity overhead lights designed for note taking during films. Videotapes were played back using two Ampex #660 2" helical-scan videotape recorders located in Roosevelt TV studio, one for cuing while the other was replaying. Commands for playing a particular speech were given by the investigator by means of a two-way intercom system.

After the playback of each speech, judges were given thirty seconds to complete their six-factor rating forms. Faculty judges then read their individual factor scores aloud and these were tabulated by all judges. The investigator announced the mean faculty score for each factor as the "best estimate of the true score" for that factor and all judges recorded that estimate for comparison with their rating. Faculty judges were then asked to discuss why they had given a particular score on a given factor and what meaning they attached to that factor. Student judges were encouraged to ask questions and make comments during discussion of the training speeches.

Since all judges were to use both 7-point and 11-point scales during the judging session, they were given training in the use of both. During the first two training speeches, all judges used Price scales containing seven discrete steps or points for each factor. In the last two training speeches, they used scales which were identical in all regards including scale end points, except that eleven equally spaced steps were available for each factor. At the end of the training period, rating forms were collected and questions answered.

Like the training portion of the pilot study, the judging phase was divided into two portions, with six speeches judged in each. Using a simple Latin square design, one half the faculty judges and one half the student judges were randomly assigned to judge the first six speeches employing the 11-point scale. The remaining faculty and student judges were given 7-point scales for the first six speeches. The speeches rated represented a random sample drawn from the recordings of pretest and posttest speeches scheduled to be judged for the master experiment. After the first six speeches had been rated, judges left the room for coffee and doughnuts. When they returned to the judging room, the investigator had collected the completed rating forms and given 7-point scales to the faculty and students who had used 11-point scales during the first half of the period and 11-point to those who had used 7-point versions. After a reminder that judges would have to continue to guard against the three common errors of raters, the second group of six speeches was evaluated. When the last speech had been judged, all ratings were collected and the judges were asked to fill out a nineteen item questionnaire.[1] Its purpose was to determine the judges' reactions to the instrument, the training phase, and the judging phase of the pilot project. The entire session took a total of three hours and fifteen minutes.

Results

Data from three different sources were used to answer the three questions posed at the beginning of this study of the rating procedure. Responses to the questionnaire provided answers to various aspects of the first question, "Will the training period and judging session be practical and efficient?" On the whole, these responses were positive. Eighty per cent of the judges felt the scale "provided a meaningful measure of 'speech skill,'" while 8% felt it did not. An average of 85% said that the training session was "helpful in increasing . . . [their] ability to use the scale with precision," that the lecture-discussion method seemed appropriate, and that the session was long enough "to become familiar with the [Price] 6-factor rating scale." When judges were given the forced choice question, "Which of the two scale lengths, 7 point or 11 point, permitted you to make more precise judgments?" 75% of the instructors chose the 11-point version. However, only 35% of the student judges picked the 11-point length. This difference was in keeping with Guilford's theory that judges with richer background and experience can effectively use rating scales of longer length than less experienced judges.[1]

---

[1] See Appendix D for the complete text and responses.

Although forty per cent of the judges indicated that the picture and sound quality were "Mildly Distracting," none felt the quality was "so poor that it seemed to hinder the precision of . . . [their] judgments." Moreover the high coefficients of reliability, reported later, suggested that the technical quality of the playback did not introduce substantial error, even though it was decidedly below commercial broadcasting standards. It is possible, although unlikely, that the technical quality may have introduced some bias to the judging, and therefore reduced validity.

One final question regarding the conditions for judging showed that fully forty-six per cent felt they were "getting so tired and/or bored toward the end that . . . [their] evaluations of the last speeches were less precise than earlier evaluations." Even though reliability estimates reproduced in Table I failed to bear out this feeling of reduced precision, the response was viewed as a danger signal and care was taken to keep judging sessions in the major experiment under two and one-half hours.

Several answers to open ended questions indicated dissatisfaction with the labeling of one of the six factors, in terms of the sub-factors which loaded on that factor. The factor label challenged was "Personality," factor #4, comprising personal appearance, posture, poise, and fluency. It was decided that the alternative title "Poise," one of the sub-factors under the fourth factor, would be suggested if similar challenges occurred during the basic experiment's training session.

---

[1] Joy P. Guilford, Psychometric Methods (2d ed.; New York: McGraw-Hill Book Company, Inc., 1954), pp. 289-96.

The second question posed for this investigation was, "Will the precision of measurement with the Price speech performance rating scale be increased by expanding scale lengths on individual factors from seven intervals per item to eleven intervals?" Treatments were balanced by using a simple Latin square design. Since "the reliability of any set of measurements," as Guilford points out, "is the proportion of their variance that is true variance," it was decided that reliability estimates would provide a convenient measure of relative precision.[1] Coefficients of reliability were computed for each cell of the Latin square by means of the Ebel "intra-class method."[2] For each of the four groups of judges, Table 1 shows resulting coefficients for scores on each of the six factors and total weighted scores.

Hotelling's special t test formula for significance of difference between coefficients of correlation which are themselves intercorrelated was used to determine whether the reliability coefficients for judgments using the 11-point scales were significantly different from those using the 7-point scales.[3]

---

[1] Joy P. Guilford, _Fundamental Statistics in Psychology and Education_ (New York: McGraw-Hill Book Company, Inc., 1965), p. 439.

[2] Robert L. Ebel, "Estimation of the Reliability of Ratings," _Psychometrika_, XVI (December, 1951), 407-24.

[3] Formula given in Guilford, _Fundamental Statistics_, pp. 190-91:

$$t_{d_r} = (r_{12} - r_{13}) \sqrt{\frac{(N - 3)\ (1 + r_{23})}{2\ (1-r^2_{23} - r^2_{12} - r^2_{13} + 2r_{23}r_{12}r_{13})}}$$

In this formula, r_{23} represents the inter-correlation of the two coefficients of correlation, N the number of subjects in the samples providing these coefficients.

TABLE 1

EBEL INTRA-CLASS RELIABILITY ESTIMATES

Faculty Judges				Student Judges			
Factor[a]	Speech Number	7-point Scale Group A_F	11-point Scale Group B_F	Factor[a]	Speech Number	7-point Scale Group A_S	11-point Scale Group B_S
1	1-6	-.8764	.5553	1	1-6	.8337	.8775
2		.8043	.7552	2		.7942	.8738
3		.9277	.8950	3		.9567	.9404
4		.4374	.7345	4		.8538	.7939
5		.3839	.3670	5		.8536	.7828
6		.6400	.6888	6		.6673	.8730
Total		.7300	.8055[b]	Total		.9001	.9278[d]
		Group B_F	Group A_F			Group B_S	Group A_S
1	7-12	.9285	.7358	1	7-12	.7931	.8723
2		.7654	.8289	2		.7035	.8010
3		.7183	.6622	3		.6436	-.2460
4		.8873	.7500	4		.7992	.8468
5		.2499	.4618	5		.8550	.6270
6		.8571	.7675	6		.5138	.7327
Total		.9057[c]	.8047	Total		.8023	.8687[e]

[a] Factor labels: (1) Content; (2) Intelligibility; (3) Bodily Action; (4) Personality; (5) Language; (6) Voice.

[b] ($p < .30$, df = 6)

[c] ($p < .05$, df = 6)

[d] ($p < .25$, df = 20)

[e] ($p < .10$, df = 20)

Before conducting t tests it was decided that unless the 7-point version showed itself superior, the 11-point scale would be used in the master experiment. This preference for the 11-point scale was established for several important reasons. Theoretical statements of Guilford[1] and

---

[1] Guilford, Psychometric Methods, Ch. 2.

Becker[1] indicate that a longer scale ought to increase the reliability of measurements where judges are more homogeneous in background, more experienced in judging the phenomena to be measured, more highly motivated, and better trained in the use of the instrument. (1) Judges selected for the major experiment were more homogeneous in terms of speech background than the pilot judges. While the major interest of many of the latter judges was theater or speech science, the prime concern and training of the major experiment's judges was public address. (2) The judges of the major experiment were more highly qualified in terms of experience in evaluating speech performances than were the pilot judges. The 11-point scale was more appropriate for them than for the pilot judges. (3) Judges in the major experiment were also more highly motivated. Unlike the pilot study judges, they were paid from $1.75 to $5.00 per hour for their services and were willing to spend twenty-five hours judging. (4) Experiment judges received nearly twice the amount of training in the use of the Price scale as their counterparts.

Four Hotelling t tests were computed using Ebel reliability coefficients based on total weighted scores. In order to show any slight difference existing between groups, to reduce the possibility of a Type II error, the decision was made to report even insignificant alpha levels. As Table I shows, the 11-point scale showed superior reliability in the two tests with ratings from twenty student judges, at the .25 and .10 levels respectively. Results for the six faculty judges showed that one group of judges was superior to the other no matter which scale length it used. This difference was significant at the .30 level when the better group used the 11-point scale, and .05 level when it used the 7-point version. The instability of ratings of three-judge groups and probable difference of variability between the first six speeches and the last six prohibit drawing conclusions favoring either scale length on the basis of these results. The fact that student judges, theoretically less able to use the longer scale, were able to demonstrate slightly more reliable measurement using the 11-point scale could be taken as evidence favoring the longer scale. On the other hand, this pilot study provides little evidence favoring the 7-point version. Therefore, because of the presumption established above, the decision was made to use the 11-point version of the Price scale in the master experiment.

---

[1] Samuel L. Becker (telephone interview, ca. February 28, 1968).

The third basic question, "Will the combined number of student and faculty judges planned for the major experiment give ratings of sufficiently high reliability?" is answered by the coefficients of reliability reported earlier. Although the highest possible reliability was sought, the level of .80 was chosen as the lowest acceptable level. Evidence from this exploratory study suggested that figure would be attained easily by pooling ratings of the major experiment's four student and five faculty judges. In the pilot study, groups of ten student judges using the 11-point scale evaluated six speeches with estimated reliabilities of .93 and .87. Groups of three faculty judges achieved reliabilities of .81 and .80 respectively, using the 11-point scale. These results also provided further evidence for the effectiveness of the training and judging sessions planned for the major experiment.

Summary

The first pilot study reported provided a test of the procedure planned for videotaping and replaying student performances for self-evaluation. Student post-test questionnaires and Instructional Broadcasting personnel conferences were used to answer several questions posed at the beginning of the study. On the basis of that information, the following answers may be stated: (1) The procedure was found to be feasible in terms of television facilities and staff available. (2) Students were enthusiastic about the value of the videotape replay activity. (3) The entire procedure did not take more time than either students or the instructor were willing to spend in terms of its apparent value. (4) Among suggestions for procedural improvements was the need to improve the training and checking of recording-playback technicians.

Two years later, the second pilot study was undertaken for the purpose of testing the instrument and procedure for measuring the experiment's dependent variable. Speeches by fundamentals of speech students videotaped in class were evaluated. Several sources of information were used to answer the three questions posed: (1) Questionnaire responses of judges indicated that, on the whole, the training period and judging session were practical and efficient. However there were indications of problems in Price's factor labels and in the length of judging sessions. (2) Reliability coefficients of faculty and student judges' ratings using both 7-point and 11-point versions of the Price scale were analyzed. Based on a simple Latin square design, this analysis provided moderate evidence favoring the 11-point version. These results, along with several theoretical considerations, led to the decision to use the 11-point form of the scale in the major experiment. (3) High reliability coefficients on ratings by faculty and student judges suggested that the number of judges selected for the major experiment would provide highly reliable judgments. This pilot study demonstrated that the training period and judging sessions planned would provide an efficient method for measuring speech skill.

CHAPTER IV

RESEARCH PROCEDURE

Elements of Design

Variables

The independent variable of greatest interest in this experimental study was the level of feedback, operationalized in three treatments: (1) videotape replay of two performances plus conventional class-instructor criticism on all performances; (2) audiotape replay of two performances plus conventional criticism on all performances; (3) "Hawthorne control" activity[1] plus conventional criticism on all performances. As pointed out in Chapter I, these three treatments represent three levels of feedback completeness. These three levels of feedback offered an appropriate and pragmatic test of the general hypothesis, "The greater the completeness and accuracy of student speech performance feedback, the greater the degree of speech skill a student will later exhibit."

The use of students from classes taught by two instructors permitted the investigation of the hypothesis as it applies to different types of recitation section teachers. The two instructors whose sections were used for the experiment represent two basically dissimilar backgrounds and sets of interests. The investigator served as section instructor for half of the subjects and TV lecturer for all of them. His major training, five years of college teaching, and research activities had been in public address and group communication. The other instructor, responsible for the other four sections, had had extensive training and experience in theater. During his five years of teaching in high school and college, he had directed a number of plays and musical productions. Although their backgrounds were dissimilar, the two instructors demonstrated agreement on the essential aspects and criteria of the course. A post-hoc comparison of their rating of the relative importance of each Price scale factor to speech skill showed a correlation of .86.

---

[1] A special session requiring subjects to view and evaluate three videotaped speeches.

Measurement of the dependent variable, speech skill, was planned for the first or "pretest" speech given by each subject and for the final or "post-test" speech. This added a third dimension to the basic design, resulting in a "3 X 2 X 2" design.[1] Because each treatment about which inferences were to be drawn was included, the design may also be designated a "fixed effects model."[2]

---

[1] B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw-Hill Book Company, 1962), pp. 337-49.

[2] William L. Hays, Statistics for Psychologists (New York: Holt, Rinehart and Winston, 1963), pp. 471-89.

Measurement Instruments

Measurement of the subjects' final speech performances in terms of the dependent variable provided a direct test of the hypothesis within the context of the experimental situation. To the extent that all significant variables were controlled or randomized, it may be assumed that any differences between group post-test performance scores was caused by differences between the three levels of feedback, differences between the two instructors, differences between the two trials, or interactions of the variables.

As pointed out in Chapter II, the instrument used to quantify the dependent variable was William Price's speech performance rating scale.[1] Nine judges, five public address instructors and four advanced speech students, employed this scale in evaluating videotapes of 216 performances by the subjects. Several subsidiary devices were also used as indirect measures of speech skill, such as total speech points and final course points for each subject, recitation class attendance, records of student-instructor conferences, and number of advanced speech courses taken by subjects after their experience in the course.[2]

---

[1] See Appendix A for the complete text.

[2] The value of collecting a variety of measures of the dependent variable is discussed in Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally and Company, 1963), p. 33.

Subjects

Subjects consisted of students enrolled in seven sections of Eastern Michigan University's three-hour fundamentals of speech course, Speech 124, Spring Semester, 1967. These classes met on Wednesdays and Fridays at 9:00 A.M., 10:00 A.M., 11:00 A.M., 12:00 noon, 1:00 P.M., 2:00 P.M., and 3:00 P.M. respectively. The videotaped lecture was presented in a large lecture hall for everyone at 10:00 A.M. each Monday. All experimental sections were scheduled for Quirk #94, a specially equipped audio-visual room seating forty-five. The eighth section of Speech 124, meeting at 10:00 A.M. in Quirk #96, was excluded from the experiment because of insufficient equipment and staff to record speeches from two classes at the same time. Because of differences in popularity of certain hours and possibly instructors, it was assumed that systematic differences existed between the seven experimental sections.

Although students at Eastern Michigan University come from widely diversified socio-economic backgrounds, the majority are from middle class families living in southeastern Michigan. Approximately 50% of the university's students plan to become teachers.[1] The average S.A.T. verbal scores of the 1967 freshmen class were 433.3 for men and 437.0 for women; math scores were 493.6 for men and 444.2 for women.[2] Since all students are required to take either the two hour fundamentals of speech course, Speech 121, or the three hour version, Speech 124, it may be assumed that the above description of the student body is reflected in the seven sections involved in this experiment.

---

[1] James Cass and Max Birnbaum, *Comparative Guide to American Colleges* (New York: Harper & Row, 1965), pp. 173-74.

[2] Robert Strauss, Assistant for Statistical Analysis and Reports, Office of Admissions, Eastern Michigan University (telephone interview, August 16, 1968).

Registration figures for experimental subjects showed that 90% of the subjects were freshmen, 6% sophomores, 3% juniors, and 1% seniors. Subject pretest questionnaire responses indicated that only 17% were twenty years old or older. The ratio of women to men was 55% to 45%, women outnumbering men. Forty-two per cent indicated that they had had a high school course involving study and practice in public speaking.

Special Installation of Equipment Required for the Experiment

Requirements

In order to videotape and audiotape student speeches as they occurred in the classroom, it was necessary to install special equipment both inside and outside Quirk #94. Three criteria were used in the design of these special facilities: (1) the presence and operation of the equipment should provide minimum distraction to the class; (2) the installation should provide maximum quality video and audio for recordings; (3) the cost should be kept at a minimum. The first criterion, minimal distraction, was considered most important and most easily controlled given the recording facilities and personnel available from Eastern Michigan University's Office of Instructional Broadcasting. Decisions on design and installation of special facilities were made in consultation with the Coordinator, LaVern

Weber, and staff engineers.

Classroom Installation

The most difficult installation problem was the unobtrusive location and operation of the television camera. The small G.E. #TE-6 industrial vidicon camera was attached to a wall mounting bracket at the back of the classroom, seven feet off the floor and immediately above the large air conditioning unit. It was operated from another room by means of an R.C.A. #PT 100 S electrical pan-and-tilt unit using two electric motors mounted under the television camera. One permitted the operator to tilt the camera up or down, the other let him pan the camera to the left or right.[1] A large slotted piece of plywood was painted to match the air conditioner and attached in front of the camera to mask it from the speaker and audience.[2]

During the test of the installation, a 70mm lens was chosen for the camera, permitting a "waist-shot" of speakers standing at the desk lectern. Although this shot excluded hand movement made at desk height, it showed gestures made on or above the lectern and facial expressions in detail.[3]

Above the lectern, an Electro-Voice #664 unidirectional dynamic microphone was suspended from the ceiling by an adjustable "goose-neck" mount.[4] On the wall behind the lectern, special medium-brown curtains were hung for contrast to prevent the white wall from "washing out" the facial expressions of the speakers. Normal overhead florescent fixtures provided sufficient light when the camera f-stop opening was set wide open. A large portable blackboard was placed to one side of the lectern, so that speakers could present visuals without walking out of camera range to the side wall blackboards. The entire installation of special equipment was completed by the first day of class and was held constant throughout the semester. It was assumed that the students accepted this equipment, along with a television receiver and retractable film screen, as part of the standard equipment of the speech classroom.

---

[1] See Appendix J, picture 4.

[2] See Appendix J, picture 2.

[3] See Appendix J, picture 9.

[4] See Appendix J, pictures 1 and 2.

External Facilities

Control of the video and audio signals was assigned to an operator positioned in the Quirk "fan room" immediately above the classroom.[1] Using a small television monitor, he could see the picture taken by the camera and control camera movement by electrical pan-and-tilt mechanism. By this means he could keep the speaker centered in the picture, although he was instructed not to follow small movement or swaying. The second part of his job required that he regulate audio volume to adhere to the basic level and dynamic range of the recording equipment used, while trying to maintain the relative difference between soft and loud speakers. He also used two Wallensak #1500 audiotape recorders to make individual 3" reel recordings of each audiotape group member's speech. Finally, he checked all speakers' names against the schedule and placed the names of audio group members on the box containing their tapes.

---

[1] See Appendix J, pictures 5 and 6.

From the fan room, both audio and video signals were carried by cable to the master control room at the TV studio in Roosevelt Hall.[1] The signals were recorded on one of three available Ampex #660 series 2" helical-scan videotape recorders. The operating engineer used the statements of the student introducers to check the speakers' names against the daily schedule. Using the index counter on the recorder, he was able to make out a daily log showing the beginning index numbers for each student's performance, and the number of the recording tape used. At the end of each recording day, the investigator compared the three lists of speeches compiled by the instructors, the fan room operators, and the master control room engineers. Where there was an inconsistency, the videotaped performance in question was viewed to determine the correct name of the speaker. By continual checking, accurate indexing was achieved, an important factor when speech ratings were to be tabulated in terms of listed index numbers.

Procedure for Employing the Independent Variables

This section provides a step-by-step description of the procedure followed in manipulating the independent variables while attempting to control or randomize all other variables of importance.

---

[1] See Appendix J, picture 7.

Preliminary Activities

Since the first day of class in the Spring Semester fell on a Wednesday, all students enrolled in Speech 124 met in separate sections at the seven different specified hours. As part of his introduction and orientation to the course, each of the two instructors read a dittoed letter purportedly written by John Sattler, Chairman of the Department of Speech and Dramatic Arts at Eastern Michigan University:

> This semester, the Department of Speech and Dramatic Arts is cooperating with the Office of Instructional Broadcasting of EMU in a research project. This project is sponsored by the United States government. . . . It involves the attitudes of fundamentals of speech students toward selected audio-visual techniques. . . . Each of you will be given written instructions several times this semester, and it is essential that you carry them out exactly as specified. . . . As important as this federal research is, what you do in it will have no effect on your grade in this course. . . .[1]

Instructors professed lack of knowledge regarding the nature of this research.

When the recitation sections met the following Friday, three experiment-related activities were carried out. First, dittoed instructions were passed out explaining the subjects' first speaking assignment, an ungraded three- to five-minute speech. Its topic was a contemporary problem of the speaker's choice; its purpose, to inform or persuade. Dittoed lists of current problem areas and possible topics were also distributed. Students were assigned speaker numbers, and the schedule for the four days of round one was read. Second, a twenty-item questionnaire was administered titled, "Department of Speech and Dramatic Arts: Questionnaire for Fundamentals of Speech Students: 121 & 124."[1] Its primary purpose was to learn the class standing, age, sex, and speech training of each subject. Finally, students were asked to make out their daily schedule on a separate dittoed schedule form, placing an "X" for each hour they were not free. The remainder of the hour was spent in brief student self-introductions given at the lectern and blackboard.

---

[1] See Appendix E for the complete text.

On the first Monday of the semester, all students met at 10:00 A.M. in a large lecture hall for the first televised course lecture given by the investigator. It dealt with general speech theory and gave specific suggestions to help students in their first speech round beginning that Wednesday. Students who had missed the previous Friday's class were asked to fill out the course questionnaire and pick up dittoes which had been distributed.

Before the next class meeting when the videotaping was to begin, subjects from each of the seven sections were assigned to the three treatment groups by means of stratified randomization, using the student as the unit for randomization. This meant that each section would include roughly equal numbers of each treatment group, thereby controlling for many of the variables which often cause sampling errors.[2]

---

[1] See Appendix F for the complete text.

[2] E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education (Boston: Houghton Mifflin Company, 1958), p. 74.

Three variables or "strata" appeared to be of sufficient importance to require balancing between groups. They were: (1) Age and class standing--subjects who were at least twenty years of age or who had reached their second semester of the sophomore year were placed in the "old" category; the remaining 81% of the subjects were categorized as "young." This stratum appeared necessary because of differences in motivation and work habits of different age groups.[1] (2) Sex--the female subjects, 55% of the total, were separated from the males because of the possibility of differences regarding the dependent variable.[2] (3) Speech training--those subjects who had taken a high school course which gave them theory and practice in developing speech skill were placed in the "yes" category; the remaining 58% were in the "no" group.[3] To insure that an equal number of each instructor's students would appear in each of the three treatment groups, subjects were also separated into two groups according to instructor.

Once the subjects had been separated into sixteen groups, according to age, speech training, sex, and instructor, they were numbered (1-k) within each group. A random number table was then used to assign them from each statified group to the three treatment groups, thus insuring that each treatment condition would contain an equal number of subjects from each stratum. Subject treatment group assignments, along with age, training, sex, and instructor, were entered on a master list of all 180 subjects. On the basis of this master list and the schedule of speaker performances, it was possible to prepare a schedule of speaker names, group designations, and order of speaking for each day of the recorded round. Copies of these schedules were given to all instructors and engineers involved. Personalized student instruction sheets were also prepared and given to the instructors for distribution after each day's speeches.[1]

---

[1] Anne Anastasi, Differential Psychology: Individual and Group Differences in Behavior (3d ed.; New York: The MacMillan Company, 1958), pp. 216-260.

[2] Ibid., pp. 453-505.

[3] Gordon Thomas (unpublished paper presented at the Spring Conference of the Michigan Speech Association, Detroit, 1966). Thomas found that students who have had previous training in speech begin a fundamentals course at a higher level than their classmates and maintain that advantage throughout the course.

One week after the first class meeting, round #1 was begun in all sections. Every speech during this four day speaking round was videotaped and preserved as a pretest trial for each subject. Performances of Group II (Audiotape) members were also recorded on individual 3" reels of audiotape. After announcements by the instructor, each speaker's name and speech title was read by the student chairman or instructor, and students spoke in their scheduled order. All of the day's six speeches were heard before any oral criticism was given. All students were asked to take careful notes on each speech in order to add to the instructor's comments during the fifteen- to twenty-minute critique period at the end of the hour. The day's speakers were reminded that they were required to hand in a written self-evaluation based on all available information at the beginning of the next class meeting, and that they should pick up special dittoed instructions bearing their names before leaving class. They were told that the instructor's written evaluation of their performance would not be available to them until their self-evaluation had been submitted.

---

[1] Campbell and Stanley, Designs for Research, p. 21.

Members of the three treatment groups who spoke each day received apparently similar dittoed instruction sheets bearing their names, explaining their required activities pertaining to the written self-evaluation.[1] Each member of Group I, the "Videotape group," was asked to go to Roosevelt television studio at one of two scheduled times to view the videotape replay of his speech. When he arrived at the television studio, an Instructional Broadcasting student operator or engineer asked him to fill out a special viewing form, giving his name, section, and date of his class performance, as well as the date and time of this replay. This permitted the operator to locate the subject's performance by the indexed list of speeches recorded on each day. It also permitted the operator to find the speech by a classmate of the subject, chosen from each section by the investigator as slightly better than average for that section, to be played along with the subject's own speech. Finally, any technical problems were noted and each form was saved as a record of subjects who had completed the viewing phase of the experiment. The student was then directed to a special viewing area separated from the rest of the studio, containing a 23" RCA classroom receiver and small desk. Taped to the desk was a copy of the instruction sheet the student had received, giving specific criteria for evaluating his own performance and suggesting that he take careful notes. The classmate's speech was played first to "validate" the videotape playback process, followed by the student's own speech played twice.

---

[1] See Appendix G for instruction sheets given members of treatment groups I, II, and III.

Each member of Group II, the "Audiotape group," was instructed to go to the Audio-Visual Center in the basement of the Library to listen to his performance. At the center, he was asked to fill out a special listening form giving his name, section, date of speech, as well as the date and time of this replay. He was then given the proper classmate's audiotape, the same performance as presented visually to the videotape members of his class. He was also given the 3" reel of his own performance, and the key to two small library rooms containing Wallensak #1500 audiotape recorders. Taped to the desks in these rooms were special instructions on how to operate the recorder, and a copy of the instructions given to Group II members stating the criteria for self-evaluation. These instructions served as a reminder that he should listen to his classmate's performance once, before listening to his own twice. When he completed listening and note-taking, he returned the tapes and key to the A-V Center desk and his returning time was noted on the listening form.

Each member of Group III, the "Hawthorne control group," was asked only to write a self-evaluation "based on all the information you have about your performance. (i.e. class comments, instructor criticism, and your own feelings.)"[1] The same critique format was given here as for the other two groups. At a later point in the semester, this group participated in a "Special Speech Improvement Activity."[1]

---

[1] Ibid.

At the beginning of the next class meeting, the instructor collected all completed self-evaluations. At the end of class, he returned student outlines with his written comments to all subjects who had submitted critiques. Instructors also made notes of any problems students were having with the experimental activities and suggested solutions where possible. For example, it was occasionally reported that an audiotape was blank. In those cases, the student was asked to go back to the A-V Center the following day, and Roosevelt TV studio was instructed to make an audiotape copy of that speech from the videotape master recording.

The above procedure was repeated on each of the four days during this first round of speaking. Those subjects who failed to do one of the activities were reminded by means of personally addressed notes given them by their instructor. The vast majority of subjects, however, completed all tasks on time without prompting. All videotaped performances were saved for later rating by judges as the pretest speeches.

Activities for the next eight weeks were the same for all subjects. Speaking assignments included a delivery work-out speech, a speech to inform, and a debate. Subjects attended lectures, completed reading and short written assignments, and took a mid-semester examination.

---

[1] See page 81.

Mid-Course Activities

During the eleventh and twelfth weeks of the sixteen-week semester the fifth round of speeches, the speech to persuade, was conducted. Although no change was made in the membership of the three groups, the order of speaking was different from that of the first round. The first round procedure of recording, viewing, and writing critiques was followed. However one small change in the viewing procedure was made; this time the classmates' comparison speeches viewed by members of the videotape and audiotape groups were chosen at random from the first day's speeches. Again instructor critiques were withheld until self-critiques had been handed in, and this time the instructor critiques also contained a grade for the speech. After students had viewed their performances, videotapes were erased to be ready for recording the final round of speaking.

Following round #5, all sections met in the Quirk television studio for two rounds of speaking on television. During these performances, a one-minute orientation speech and a five-minute speech, students were instructed to look into the camera lens as they spoke. For this reason, subjects rarely saw themselves on the floor monitor.

Five days before the final or "post-test" round of speaking began, each member of Group III (Hawthorne control group) received instructors to report to the Roosevelt TV studio for a "Special Speech Improvement Activity."[1] Its publicized purpose was to help students give better final speeches. The actual purpose of this activity was to control for the "Hawthorne effect," the behavioral difference between experimental and control groups caused by differences in perceived interest and attention of the experimenters towards different groups of subjects.[1]

---

[1] See Appendix H for the complete text.

After Group III members had assembled, the course lecturer (investigator) handed out dittoed booklets containing instructions and speech evaluation forms.[2] The instructions falsely stated that research just published in Speech Monographs had demonstrated that "a beginning speech student can improve substantially by viewing a variety of video taped [sic] speeches, evaluating those speeches in terms of stated criteria, and comparing each speaker's strengths and weaknesses with his own." They were then shown videotapes of five- to ten-minute speeches by three different speakers: (1) a Speech 124 student recorded during round #5; (2) Eastern Michigan University President Sponberg in a portion of his inaugural address of the previous year; (3) Hal Holbrook portraying Mark Twain lecturing at the turn of the century. The subjects were instructed to write their evaluation of each speech according to the criteria supplied, and to think of their own speaking in terms of each of the criteria. The activity lasted approximately fifty minutes. The several members of group III who did not attend were invited to make up the exercise two days later. Four days after this activity was completed, the final round of speaking was begun.

---

[1] Fritz Roethlisberger and W. J. Dickson, Management and the Worker (Cambridge: Harvard University Press, 1939).

[2] See Appendix I for the complete text.

Final Round Activities

The final speech was scheduled as a 4-6 minute "speech on a contemporary problem." Dittoed instructions for this round were identical to those for the first speech, except that the final speech constituted one-fifth of the total course grade rather than being ungraded. All students were again required to hand in written self-evaluations before receiving the instructor's comments or grade. However for this round it was announced that Instructional Broadcasting would attempt to videotape all speeches to determine the feasibility of recording on such a massive scale. Students were told that because of scheduling difficulties, no one would be permitted to view or hear the recordings of his performances.

At the final examination an objective type questionnaire was given to determine subjects' overt reactions to various aspects of the experiment.[1] So no subject would think his responses would influence his course grade, the signed questionnaires were administered and collected by the Coordinator of Instructional Broadcasting.

During the semester, 443 speeches given on sixteen different days had been recorded on videotape, while only two had not. Of the 180 subjects originally assigned to treatment groups, 123 completed all ten experimental activities. Approximately the same number of subjects from each treatment cell successfully completed all experimental tasks. However slightly more of instructor A's subjects completed all tasks than instructor B's. In order to make each of the six treatment-instructor cells equal, fifteen subjects were randomly discarded, leaving a total of 108 subjects, with 54 per instructor cell, 36 per treatment cell, and 18 per instructor-treatment cell. This number of subjects was considered appropriate to permit statistical analysis of sufficient power to discover any real differences between treatment groups.

---

[1] See Appendix K for the complete text.

Procedure for Measuring the Dependent Variable

Selection of Judges

The primary means of quantifying the dependent variable was by employing trained judges to rate videotaped pretest and post-test speeches using the 11-point version of the Price speech performance rating scale.[1] Nine judges were selected for the rating procedure, five faculty and four students. The following criteria were used in making the selection of faculty and student judges: (1) amount of experience teaching or taking courses in public address and group communication; (2) level of motivation for judging; (3) experience in judging or performing in forensic contests; (4) basic scholarship level in speech.

All five faculty judges had taught fundamentals of speech courses for at least four years, with the average 6.3 years. Each agreed to spend twenty-five hours evaluating videotaped speeches by beginning students, and received four or five dollars per hour, depending on his background. Each had judged during several seasons of Michigan Forensic League competition. Four had master's degrees in speech while the senior member had completed his doctorate. Four of the five were men, one a woman.

---

[1] See Appendix A.

The four student judges had an average of thirty-nine semester hours in speech and drama courses at Eastern Michigan University. Of these, an average of twenty-three hours was in public address and group communication. Since all four planned to teach speech in college or high school and saw this judging experience as valuable training, each appeared highly motivated. Each student judge received $1.75 per hour. Three of the student judges had participated in collegiate forensic contests. These judges consisted of one sophomore, one junior, one senior, and one graduate student. Their mean overall grade point average was 3.15, or slightly above a "B." Three of the four were women.

Training Period

The procedure for the two-hour training period used in this study was similar to that tested in the pilot study.[1] The purpose of this period was to increase the reliability of measurements given by judges using the Price scale. Specifically, this was accomplished by helping the judges establish: (1) an understanding of the common errors made by all raters; (2) anchoring of the scale extremities in terms of the performances to be judged; (3) a common understanding of the meaning of each factor on the Price scale.

---

[1] See Chapter III.

The first phase of the training period consisted of a ten minute lecture with projected graphs dealing with the three most common errors of judging--central tendency, halo effect, and leniency effect. Suggestions were made for correcting each error.

The second and more important phase consisted of judging six videotaped performances, with discussion by faculty members following each. The six performances had been selected by the investigator as representative of the best, the worst, and the average speeches given by subjects who had failed to complete all experimental tasks. These performances were selected to help the judges anchor the 11 steps on each scale factor in terms of the population of subjects to be judged.

Faculty members discussed each speech, factor by factor, to help all the judges come to a common understanding of the meaning and characteristics of each of the six scale factors. After each speech the investigator asked the faculty judges to announce their scores on each factor in turn. The investigator tallied the scores and announced the mean faculty score on each factor as the "best estimate of the true score." Any discussion of the factor which seemed appropriate took place at that time. Judges were told not to be concerned if their scores were consistently above or below the best estimate score, since the Ebel estimate of reliability, which was to be computed on the ratings, adjusted for consistent inter-rater differences.

The only factor label challenged was #4, "Personality," as it had been in the pilot study.[1] The judges accepted the suggestion that "Poise" was a better designation of that factor in terms of the sub-factors. On factor #5, "Language," it was decided that the major criterion for judging would be clarity and precision of language rather than its vivid and imaginative use. At the end of this six-speech training session, judges expressed confidence in their ability to use the Price scale effectively. Following a fifteen-minute break, they began rating videotaped speeches from the major experiment.

---

[1] See Appendix A.

Judging Sessions

Speeches were judged in groups of twelve, with thirty seconds between speeches for completion of ratings. All judging sessions except the first consisted of rating two groups of twelve speeches, with coffee and doughnuts after the first twelve. These sessions lasted an average of two and one half hours and were conducted on Tuesday, Friday, and Sunday evenings, as well as on Saturday mornings. During the three weeks of judging, a total of 216 different speeches were rated.

The possibility of an "order effect" in judging, the evaluation error caused by the order of presentation of performances, was controlled by arranging speeches on the basis of stratified randomization. This method not only controls for the effect of one speech upon the judging of another within any session, it controls for differences in judge morale, health, and fatigue from session to session. The 3 X 2 X 2 design for the three treatments, two instructors, and pre-post speeches meant that twelve different cells existed, each containing eighteen speech performances.

To accomplish random assignment of the speeches on the basis of membership in these twelve cells or strata, eighteen random orderings of the numbers 1-12 were drawn and recorded. Next to each of those randomly ordered cell numbers, a name and tape index number were noted, randomly drawn from the speeches of that cell. Therefore any group of twelve speeches judged during the first or second half of a session contained a random ordering of representatives from each of the twelve cells. In the several instances where one speaker's pretest and post-test speeches were scheduled for the same judging session, one in the first half and one in the second, either was randomly switched to another judging period. To maintain the "double-blind" feature of the judging, no suggestion was made to the judges that half these speeches had been given at the beginning of the semester and half at the end, or that they represented different treatments or different instructors.[1]

Since the engineer was required to play a specified ordering of 216 speeches from the more than 300 located on nineteen 2400' reels of videotape, he used two recorders for playback. This permitted him to cue up one performance while another was being played into Quirk #94. The investigator was in the judging room, checking the scheduled playing order against the actual speeches played, and adjusting the picture and volume when necessary. The several times when the wrong speech was played because of inaccuracy of the recorder index counter, the investigator quickly changed channels and called the engineer by intercom to correct the problem.

---

[1] Campbell and Stanley, Designs for Research, p. 14.

For a warm-up at the beginning of each of the first five sessions, the investigator had a performance played which had been judged in the training period or a previous judging session. Judges rated the speech on practice forms and compared their ratings with mean scores given during the previous rating of the speech. After the warm-up, judges were given booklets of twelve rating scales coded for the session number. They were asked to write their names and the numbers of the speeches to be judged on each form. When all forms had been prepared, the investigator called the engineer to play the first tape.

Occasionally one or more judges had to miss a judging session and a make-up was scheduled. Since the stratified randomized order of playback kept such make-up judging from having any systematic effect on ratings, any error introduced by these special sessions could only have a conservative effect on the results by increasing error variance. The make-up sessions were conducted in the same way as the regular judging periods.

To provide data for the computation of test-retest reliability, the group of twelve speeches judged in the beginning of the third session was again judged two weeks later in a different order. This reliability estimate was planned as a check on the Ebel method of estimating reliability used for the 216 speech measurements. After the retest judging had been completed, a short questionnaire was administered to learn the judges' feelings about the experiment, the Price scale, and the procedure for using it.[1]

---

[1] See Appendix L for complete text and responses.

Summary of Procedure

The three independent variables investigated in this experimental study were feedback mechanism, instructor, and repeated measurement. The feedback mechanism was operationalized in three treatments representing three levels of feedback: (1) videotape replay of two performances plus conventional criticism on all performances; (2) audiotape replay of two performances plus conventional criticism on all performances; (3) "Hawthorne control" activity plus conventional criticism on all performances. The two instructors represented dissimilar interests and backgrounds: (1) public address and group communication; (2) theater. Measurements of two trials for each subject were taken: (1) pretest performance; (2) post-test performance. These three levels of feedback, two instructors, and two measurements of the dependent variable represented a 3 X 2 X 2 fixed effects model.

To minimize the distraction caused by videotaping class speeches, a small television camera was installed on the back wall of the classroom with a face board hiding it from view. A technician located in another room controlled camera movement with a remote pan-and-tilt device. He also monitored and recorded the audio signal picked up by the microphone suspended above the lectern. Both audio and video signals were then carried by cable to the master control room where they were recorded on videotape. All speeches were carefully checked and indexed.

Subjects were Eastern Michigan University students enrolled in seven sections of Speech 124, Fundamentals of Speech, during the Spring Semester, 1967. They were stratified by age, sex, and speech background, and randomly assigned to the three feedback conditions. Using the subject as the unit of randomization meant that each of the seven sections contained roughly equal numbers of subjects from each treatment group. Subjects were not informed of their experimental treatment until after they had completed their first speech assignment, a three- to five-minute speech on a contemporary problem. All speeches given during this round were videotaped and saved as pretest speeches.

Following their speeches, members of Group I (videotape) were instructed to go to the television studio at a specified time to view the videotape of a classmate's speech once and their own twice. They were required to hand in a written self-evaluation at the next class meeting. Members of Group II (audiotape) were asked to go to the Audio-Visual Center to listen to an audiotape recording of a classmate's performance once and their own twice. Like the videotape group, they were required to hand in a written self-evaluation. Members of Group III (Hawthorne control) were simply asked to evaluate their performance on the basis of information they had at their disposal and to hand in that analysis at the next class meeting. No student received his instructor's critique until after he had handed in his self-evaluation.

During the fifth round, the speech to persuade, the above procedure was repeated for all subjects, although the speaking order was changed. Before the final speech round, members of Group III, the "Hawthorne control" group, were instructed to attend a "Special Speech Improvement Activity" in the television studio. After being told that this activity would improve their speaking, they viewed and evaluated videotaped speeches of three speakers.

During the final round of speaking, a four- to six-minute speech on a contemporary problem, performances of all subjects were again videotaped and preserved as post-test speeches. During this round all subjects were required to submit written self-evaluations, although none were permitted to view their performances.

Nine judges were selected to rate videotapes of subject pretest and post-test performances, using the Price scale. The five faculty members and four speech students comprising this panel were chosen on the basis of background in public address, motivation for judging, experience in judging, and scholarship in speech.

Judges were trained in the use of the measurement instrument by means of a brief illustrated lecture and practice ratings and discussions of six representative speeches. Subject performances were stratified on the three levels of feedback, two instructors, and two measurements, and randomly ordered in groups of twelve. Judging of the 216 speeches was accomplished in ten rating sessions during a three-week period.

CHAPTER V

RESULTS

This chapter describes the establishment of reliability and validity of measurements of the dependent variable. A statistical model for analyzing the data is presented, along with results of that analysis. Findings based on analyses of subsidiary measurements are also discussed briefly.

Reliability of Measurements

An investigation of the reliability of judges' measurements of the dependent variable is essential before discussing any results based on those measurements. As Guilford points out, "Conclusions derived from statistical results might differ considerably depending upon how reliable we know the measurements to be. Thus, the matter of reliability merits considerable attention."[1]

The Ebel "intra-class" procedure for estimating reliability of performance ratings was employed in this study.[2] This procedure permits the total variance of a series of ratings by several judges to be divided into "three components, attributable to pupils, raters, and error. . . . Thus it is possible, if desired, to remove the 'between-raters' variance from the error term."[1] By removing the variance caused by between rater differences which distorts reliability estimates, recognition is given the fact that a rater who consistently rates above or below the mean rating may nonetheless add substantially to the actual reliability of the combined measurements.

---

[1] Joy P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill Book Company, 1965), p. 238.

[2] Robert L. Ebel, "Estimation of the Reliability of Ratings," Psychometrika, XVI (December, 1951), 407-24.

Table 2 shows the results of the Ebel intra-class reliability estimates of Price scale total weighted scores for 216 speeches given by nine trained judges. Reliability coefficients are presented for the average single judge as well as for all nine judges for each of the six factors on the Price scale and for the weighted total.

TABLE 2

EBEL ESTIMATE OF RELIABILITY OF RATINGS ON 216 SPEECHES

Factor	One Judge	Nine Judges	9 Judges: Estimate Rounded
#1 Content	.4685	.8880	.89
#2 Intelligibility	.5237	.9082	.91
#3 Bodily Action	.6051	.9324	.93
#4 Personality (Poise)	.6117	.9341	.93
#5 Language	.4331	.8730	.87
#6 Voice	.5228	.9079	.91
Weighted Total$_{\text{Price}}$	.6577	.9453	.95

---

[1] Ibid., p. 411.

With nine judges, it was possible that the ratings of one could have been so unreliable as to reduce the overall reliability of the pooled judgments. This contingency was investigated by selectively deleting one judge at a time from ratings of twelve speeches and computing reliability coefficients for each remaining group of eight judges. Results of this procedure, shown in Table 3, indicated that no single judge, faculty or student, reduced the overall reliability of the ratings.

TABLE 3

EBEL ESTIMATES OF RELIABILITY OF RATINGS ON TWELVE SPEECHES WITH ONE JUDGE OMITTED

	Omitted Judge									
Factor	None	1 (Fac)	2 (Fac)	3 (St)	4 (St)	5 (Fac)	6 (Fac)	7 (St)	8 (Fac)	9 (St)
Content	.82	.78	.80	.81	.79	.78	.82	.78	.84	.78
Intell.	.89	.88	.88	.90	.87	.86	.89	.87	.87	.88
Bod. A.	.95	.94	.93	.94	.95	.94	.95	.94	.94	.94
Person.	.96	.95	.95	.96	.95	.95	.95	.95	.95	.95
Lang.	.93	.92	.93	.92	.93	.92	.94	.92	.93	.92
Voice	.97	.96	.96	.97	.96	.96	.96	.96	.96	.96
Tot$_W$	.96	.96	.96	.96	.96	.96	.96	.96	.96	.96

For a comparison of reliability of ratings by student judges to those by faculty judges, separate Ebel estimates were computed for student and faculty judgments of the 216 speeches. To achieve equality in the number of judges per group, one faculty judge, number six, was randomly selected and his ratings were omitted from these comparison groups of judgments. The Ebel reliability coefficient for four student judges' total weighted scores on 216 speeches was

.89; the coefficient for four faculty judges' scores was .86. A Hotelling t test showed no significant difference between rating reliability of four faculty and four student judges.[1] Although it is possible that validity of student judgments was lower than for faculty judgments, the intercorrelation of those two sets of total weighted scores (.91) argues against this contingency.

As a check on the magnitude of reliability estimates computed by the Ebel procedure, test-retest reliability coefficients were also computed. Data for these correlations were acquired by requiring the judges to rate twelve speeches played in two different random orders during sessions separated by two weeks. It was assumed that the judges would not be able to remember their original ratings because in the interim they had rated 167 other speeches. The test-retest reliability of pooled total weighted scores was .93.

Table 4 shows each judge's test-retest reliabilities for total weighted scores listed in rank order.

A Hotelling t test between test-retest reliability coefficients for ratings by four student judges (.91) and four faculty judges (.94) shows no significant difference between student and faculty judgments. Intercorrelation between these two sets of judgments was again high, .92.

Validity of Measurements

Because of the absence of other precise measures of speech skill, no direct attempt was made to establish the validity of measurements of the dependent variable in this study.

---

[1] Guilford, Statistics, pp. 190-91.

TABLE 4

RANK ORDER OF TEST-RETEST RELIABILITY COEFFICIENTS FOR EACH JUDGE ON TWELVE SPEECHES

Rank	Classification	Reliability	Judge Number
1	Faculty	.94	2
2	Student	.91	3
3	Faculty	.88	5
4	Student	.87	7
5	Student	.84	9
6	Faculty	.77	1
7	Faculty	.74	6
8	Faculty	.65	8
9	Student	.55	4

The developer of the measurement instrument, William Price, used several methods of validation during his research with the scale.[1] His multiple regression analysis to determine factor weights led to the finding of a .95 coefficient of multiple determination. "Thus 95% of the variance of the criterion measure of speaking ability, the sum of the five speeches [graded by the class professor], is accounted for by whatever is measured by the six items of the rating scale."[2] This provides evidence for a form of "factorial validity."[3] However, as Price admits, the use of one faculty member's grades of five speeches provides a questionable criterion measure.[1]

---

[1] William K. Price, "The University of Wisconsin Speech Attainment Test" (unpublished Ph.D. dissertation, University of Wisconsin, 1964), pp. 205-309.

[2] *Ibid*., p. 267.

[3] Guilford, *Statistics*, pp. 470-71.

In spite of the fact that there is no appropriate criterion measure to compute a meaningful coefficient of validity for the measurements in this study, their validity may be accepted on the basis of the following: (1) the high reliability of measurements using this instrument, indicating agreement of experienced judges whose evaluations of speech skill may be assumed to have relatively high validity; (2) one of the necessary elements for high validity, namely high reliability (although high reliability does not insure high valdity); (3) the high correlation among individual judges' weights for scale factors; (4) the validating procedure used by Price.

Statistical Manipulation of Major Data

A 3 X 2 X 2 fixed effects design was used for analysis of data provided by the nine judges. To appropriately employ analysis of variance, several assumptions had to be met. According to Lindquist, these are:

> 1) All treatment groups were originally drawn at random from the same parent population.
>
> After administration of the treatments, each group may then be regarded as a simple random sample from a different (hypothetical) treatment population.
>
> 2) The variance (σ^2) of the criterion measures is the same for each of these treatment populations.
>
> 3) The distribution of criterion measures for each treatment population is normal.
>
> 4) The mean of the criterion measures is the same for each treatment population (the null hypothesis).[2]

---

[1] William K. Price (Personal interview, Central States Speech Convention, Chicago, 1968).

[2] E. F. Lindquist, Design and Analysis of Experiments

Lindquist's first requirement, random assignment of treatment groups drawn from one parent population, was satisfied in this study through the use of the stratified randomization technique described in Chapter IV.

The second requirement, homogeneity of variance, was tested by means of Hartley's test using the sampling distribution of the F_{max} statistic.[1] For the three-way analysis of variance with one repeated measure model used, two F_{max} ratios must be computed, one for the "subjects within groups" and one for "C X subjects within groups" representing the partition of within cell variation.[2] Since the analysis of variance was conducted eight times, once for each of the six factors and two weighted totals, a total of sixteen Hartley tests were computed in all. Results of these tests are given in Table 5.

For an alpha level of .05, an F_{max} ratio higher than the critical value would be expected to occur by chance one time in twenty. For this reason, the results here failed to provide strong evidence for the absence of

---

in Psychology and Education (Boston: Houghton Mifflin Company, 1958), p. 73.

[1] Although it is less precise than the widely used Bartlett's test, Winer recommends it over the latter for most purposes: "There is no need . . . for a high degree of sensitivity in such tests [for homogeneity of variance], because F tests are robust with respect to departures from homogeneity of variance." The formula for Hartley's test is as follows:

$$F_{max} = \frac{\text{largest of k treatment variances}}{\text{smallest of k treatment variances}}.$$

See B. J. Winer, _Statistical Principles in Experimental Design_ (New York: McGraw-Hill Book Co., Inc., 1962), p. 93.

[2] _Ibid_., p. 339.

TABLE 5

RESULTS OF HARTLEY TESTS FOR HOMOGENEITY OF VARIANCE OF DATA

Factor	#1	#2	#3	#4	#5	#6	Tot_P	Tot_J
Subj W Grp (Error Betw)	1.941	2.018	4.396[a]	2.420	1.739	3.093	2.285	2.379
C X Subj W Grp (Error Within)	3.164	2.625	3.201	2.878	3.649	2.298	2.014	1.864

[a] (Critical Value $F_{max} = 4.30$ for $p < .05$, df = 6,17.)

homogeneity of variance. Therefore this assumption was apparently met.

The assumption of normally distributed criterion measures for each treatment group has little real importance for using F tests, unless distributions are markedly peaked or flat. As Lindquist states, "In general the F-distribution seems so insensitive to the form of the distribution of criterion measure that it hardly seems worthwhile to apply any statistical test to the data to detect non-normality, even though such tests are available."[1] In order to determine whether the speech rating data was so non-normal as to prohibit the use of analysis of variance, randomly selected data from several treatment groups were graphed. Inspection of these graphs indicated that none of the distributions were markedly peaked or flat, satisfying the normal distribution assumption.

---

[1] In support of this statement, he presents the results of Norton's studies of the effects of distribution non-normality and heterogeneity of variance upon F distribution alpha levels, determined empirically and read from the F table. See Lindquist, Design and Analysis, p. 86.

The fourth assumption, equal mean criterion measures for all groups, is the null hypothesis to be tested by the analysis.

The theoretical model and computational procedures for this study were provided by Winer.[1] He designates the fixed effects design involving three factors with repeated measurements on one of the three factors as "Case II." In this design, each group is observed under all levels of factor C (repeated measurements), but each group is assigned to only one combination of factors A and B (feedback treatments and instructors). Because there are repeated measures on factor C, the within cell variation or "error term" is subdivided into two parts, the $SS_{\text{subject within groups}}$ or $SS_{\text{error (between)}}$, and the $SS_{\text{C X subject within groups}}$ or $SS_{\text{error (within)}}$. The former source of variation, the "between subject error," is a measure of the extent to which the mean of a subject differs from the mean of the group in which the subject is located.[2] The latter source of variation, the "within subject error," is a measure of the extent to which the subject's gain between repeated measures (factor C) differs from mean gains of the group in which the subject is located.[3]

---

[1] Winer, Statistical Principles, pp. 337-49.

[2] $SS_{\text{subj w. groups}} = \frac{\Sigma\Sigma\Sigma P^2_{m(ij)}}{r} - \frac{\Sigma\Sigma(AB_{kj})^2}{nr}$, where $\Sigma\Sigma\Sigma P^2_{m(ij)}$ represents the sum of the squared totals for each subject, r the number of repeated measures, $\Sigma\Sigma(AB_{kj})^2$ treatment-instructor cell totals squared, and n the number of subjects per cell.

[3] $SS_{\text{C X subj W. groups}} = \Sigma\Sigma\Sigma\Sigma X^2_{ijkm} - \frac{(ABC^2_{ijk})}{n}$,

In figuring F ratios and planned comparison t tests based on this model, the $MS_{\text{error (between)}}$ provides the denominator for analysis of simple effects of A and of B as well as the A X B interaction. All F ratios and t comparisons involving simple or interaction effects for factor C use the $MS_{\text{error (within)}}$ term in the denominator.[1] It is not unusual for this $MS_{\text{error (within)}}$ term to be as little as one-tenth the size of the $MS_{\text{error (between)}}$.[2]

Several different tests of significance were planned using results of the Case II analysis. Hays points out that a small number of powerful "planned comparisons" may be computed in place of the more common F tests, whenever the theoretical framework gives strong reason to anticipate that a given group mean will be greater than that of another group:

> The F test is an "omnibus test" of all possible comparisons to be made among a particular set of means in the data. Each of an independent set of these comparisons can be tested separately instead of all together, if the experimenter has definite questions to ask about his data to begin with. . . . The important

---

where $\Sigma\Sigma\Sigma\Sigma X^2_{ijkm}$ represents the sum of squares of all scores and (ABC^2_{ijk}) is the sum square total of all twelve ABC (treatment-instructor-measurement) cells.

[1] The applicability of this analytical procedure to the study reported here is shown by this study's similarity to the experiment Winer uses to illustrate computational procedures for Case II. His example is a modified version of the Meyer and Nobel study on the effect of anxiety (factor A) and muscular tension (factor B) on a learning task. Subjects were assigned to four combinations of A and B, and were measured during four blocks of trials (factor C). See Winer, Statistical Principles, pp. 341-49.

[2] Ibid., p. 315.

> thing to remember is that for each degree of freedom in the sum of squares between groups or treatments, there is a potential prior question to be asked of the data.[1]

The practical value of planned comparisons lies in the fact that their use decreases the chances of the Type II error, that of retaining a false null hypothesis, without increasing the chances of a Type I error. Therefore, where theory had predicted particular differences between treatment groups, a priori comparisons were planned for each degree of freedom available.[2] Since relationships between treatment groups were theorized, factor A (feedback treatment) was investigated using planned comparisons, both in testing for simple effects and for interaction with factor C, the repeated measure. Of the three simple effects and three interaction effects comparisons suggested, only two, the number equal to the degrees of freedom, were appropriate for testing in this manner. Theory also suggested that the post-test speeches (C_2) would show higher mean scores than the pretest speeches (C_1). Therefore this question was also tested using planned comparisons.

Specifically, the null hypotheses based on research

---

[1] William L. Hays, Statistics for Psychologists (New York: Holt, Rinehart and Winston, 1963), p. 475.

[2] Formula for planned comparisons:

$$t = \frac{\hat{\psi}}{\sqrt{\text{est. var. } (\hat{\psi})}} = \frac{M_1 - M_2}{\sqrt{MS_e \left(\frac{1}{N} + \frac{1}{N}\right)}},$$

where $\hat{\psi}$ is the estimate of some particular population comparison, M_1 and M_2 are the two group means constituting that comparison, N is the number of subjects in each compared cell, and MS_e the mean square error associated with that comparison. See Hays, Statistics, pp. 464-66.

hypotheses and tested by planned comparisons were:

(1) $A_1 \leq A_2$ (Videotape group equal to or less than Audiotape group.)

(2) $A_2 \leq A_3$ (Audiotape group equal to or less than H. control group.)

(3) $A_1C_2 \leq A_2C_2$ (Videotape post-test speech equal to or less than Audiotape post-test.)

(4) $A_2C_2 \leq A_3C_2$ (Audiotape post-test speech equal to or less than H. control post-test.)

(5) $C_2 \leq C_1$ (Post-test speeches equal to or less than pretest speeches.)

Null hypotheses involving B (instructors) were tested by means of F tests. Where these showed the existence of some systematic non-chance differences, "post-hoc" comparisons were planned. This type of test is less powerful than multiple t tests, but it is far more appropriate following a significant F test.[1] The specific post-hoc or "a posteriori" test chosen was the Newman-Keuls procedure which makes use of the studentized range statistic.[2]

Results of Analyses of Major Data

In this section, results of a total of eight Case II three-way analyses of variance with repeated measures on one variable will be summarized. In order to get one score for each of the 216 performances, the arithmetic mean of the nine judges' ratings for each Price scale factor was taken for each speech. As noted earlier, these mean ratings showed an overall reliability of .95, and median scale factor reliability of .91.

---

[1] Ibid., p. 472.

[2] Winer, Statistical Principles, pp. 77-89.

Using this set of mean factor ratings, two different sets of total weighted scores were computed for the speeches: one using weights resulting from Price's multiple regression analysis, the other using mean weights assigned by faculty judges in this experiment. The latter set of weights was determined by asking faculty judges in this study to rate the relative contribution each of the six factors makes to speech skill. The resulting weights are given in Table 6, along with the Price weights. However these two sets of factor weights cannot be thought of as comparable since they were generated in totally different ways.

TABLE 6

FACTOR WEIGHTS FOR THE PRICE SCALE: TWO SOURCES

Factor:	#1 Content	#2 Intell.	#3 B.Act.	#4 Person.	#5 Lang.	#6 Voice
Price MRA Weights (% of total)	0.54 (18%)	1.00 (33%)	0.37 (12%)	0.31 (10%)	0.26 (9%)	0.54 (18%)
Judge$_F$ Est. Weights (% of total)	1.09 (36%)	0.16 (5%)	0.38 (13%)	0.50 (17%)	0.43 (14%)	0.45 (15%)

Computation of the correlation between the two sets of factor weights showed a coefficient of -.28. A similar correlation comparing the Price factor weights with the two instructors' estimates of the relative importance of each scale factor to speech skill showed a similarly low coefficient, -.08. As reported earlier, the degree of agreement between the judge weights and the instructor weights was .89. Because of the discrepancy between the two sets of weights, Case II analyses were conducted using both sets of weights.

Results of analyses using the two different sets of factor weights led to essentially the same results, a significant treatment effect favoring videotape feedback. Total score means and standard deviations of groups involved in relationships showing significant differences are given in Table 7. Results of the three-way analysis of variance with repeated measures are shown in Table 8, while Table 9 gives the results of planned comparisons for hypothesized relationships.[1]

On post-test performances (C_2), the mean score for the videotape group (A_1) was significantly higher than the mean score for the audiotape group (A_2) ($p < .01$, df = 102). This result means that subjects who viewed videotapes of two of their class speeches demonstrated greater speech skill on final speeches than subjects who heard audiotapes of two of their class performances. While this finding was consistent with the general experimental hypothesis, the next result was not. It indicated that post-test speech scores of the audiotape group (A_2) were not significantly higher than post-test scores for the "Hawthorne control" group (A_3). These results were established by planned

---

[1] See Appendix M, Tables 14 and 15, for comparable results of three-way analysis of variance and planned comparisons using factor weights provided by the experimental judges.

TABLE 7

MEANS AND STANDARD DEVIATIONS FOR GROUPS INVOLVED IN IMPORTANT RELATIONSHIPS ON TOTAL WEIGHTED$_{\text{Price}}$ SCORES

Measurement	Pretest (C_1)		Post-test (C_2)	
Statistic	Mean	S.D.	Mean	S.D.
Videotape (A_1)	18.40	4.13	20.98	3.63
Audiotape (A_2)	17.90	3.64	19.74	4.04
H. Control (A_3)	17.45	4.15	19.32	4.17
All Groups	17.92	3.98	20.01	3.95
Video-Inst$_{\text{P.A.}}$ (A_1B_1)	17.40	3.32	19.97	3.54
Video-Inst$_{\text{Th.}}$ (A_1B_2)	19.40	4.69	21.98	3.53
Audio-Inst$_{\text{P.A.}}$ (A_2B_1)	18.88	4.19	21.20	4.26
Audio-Inst$_{\text{Th.}}$ (A_2B_2)	16.91	2.76	18.28	3.30
H. Con-Inst$_{\text{P.A.}}$ (A_3B_1)	17.48	4.23	19.58	4.09
H. Con-Inst$_{\text{Th.}}$ (A_3B_2)	17.41	4.19	19.06	4.35

comparison t tests conducted in place of F tests for simple effects on factor A, factor C, and on the AC interaction.

When pretest speech scores were combined with post-test scores to test for simple effects on factor A, levels of feedback, the videotape group was not significantly higher than the audiotape group. This was because the similarity of treatment group pretest scores masked the differences between group post-test scores, and because this analysis made use of a larger within group variance term than those involving factor C, pre-post measure.

Mean post-test performances for subjects taken without regard to treatment group or instructor demonstrated

TABLE 8

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON TOTAL WEIGHTED$_{\text{Price}}$ SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	2934.4375	107		
A (3 treatments)	63.8750	2	31.9375	(1.2116)[a]
B (2 instructors)	3.3125	1	3.3125	0.1257
AB	178.6250	2	89.3125	3.3883[b]
Subj. W. Groups	2688.6250	102	26.3591	
Within Subjects	676.6875	108		
C (2 measurements)	237.9375	1	237.9375	(56.7876)[a]
AC	6.3125	2	3.1562	(0.7533)[a]
BC	2.8750	1	2.8750	0.6862
ABC	2.1875	2	1.0937	0.2610
C X Subj.W. Groups	427.3750	102	4.1900	

[a] Planned comparison t tests were conducted in place of the F tests presented here within parentheses.

[b] ($p < .05$, df = 2,102)

TABLE 9

RESULTS OF PLANNED COMPARISONS ON TOTAL WEIGHTED$_{\text{Price}}$ SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.205
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.104
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	2.570[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	0.876
$C_2 \leq C_1$ (Post-test ≤ Pretest)	4.332[b]

[a] ($p < .01$, df ≤ 102)

[b] ($p < .001$, df ≤ 102)

significantly greater speech skill than mean pretest performances ($p < .001$, df = 102).

An arithmetic comparison of treatment subject gain scores on total weighted points indicated that subjects who received videotape feedback improved forty per cent more during the semester than subjects who received audiotape or no electronic feedback.

A slight interaction effect on total weighted scores between level of feedback and instructor (AB) was revealed by an F ratio of 3.388 ($p < .05$, df = 2, 102). This finding suggests that one instructor was relatively more effective using one type of electronic feedback procedure, while another found another procedure more beneficial. However a two-way analysis of variance conducted with pretest measures only indicated the presence of a similar interaction effect ($p < .15$, df = 2, 102) resulting from sampling error.[1] At the time of this pretest performance, none of the subjects could have known what feedback treatment he would receive during the experiment.

To compensate for this sampling error bias, subject gain scores were computed and analyzed, showing that any interaction effect between treatment and instructor probably involved only the audiotape and "Hawthorne control" groups. This indicated that the instructor with public address background was arithmetically more effective employing audiotape feedback, while the instructor with theater background was relatively more effective using videotape replay of other speakers.

---

[1] See Appendix N, Table 29 for the complete two-way analysis of variance. This analysis uncovered no other differences even remotely approaching significance.

Since mean gain scores for videotape subjects were identical under both instructors, that feedback procedure may have had a stabilizing effect on treatment-instructor interaction. However, these observations are only tentative implications of the data. Only the existence of an AB interaction was established by this analysis of data.

In order to investigate in greater detail the effect of videotape and audiotape feedback, a priori analyses of individual Price scale factor scores were carried out. Table 10 shows treatment group means and standard deviations from these procedures. Analysis of post-test scores on factor #1, Content, indicated no significant differences between treatment groups.[1] However, the arithmetic difference favoring the videotape group over the audiotape group approached significance ($t = 1.309$; $p < .01$, $df = 102$). Although the videotape group was not significantly better than the audiotape group in post-test performances, subjects taken without regard to group were significantly better in post-test performances than in pre-test, resulting in a t value of 12.396 ($p < .001$, $df = 102$).

On factor #2, Intelligibility, the arithmetic differences favoring videotape over audiotape and audiotape over control activity did not approach significance.[2] Nor was the arithmetic difference favoring the post-test performances of all subjects over pretest performances ($p < .10$, $df = 102$).

Because videotape replay was unique among treatments in providing feedback of the visual code, that procedure was expected to provide a significant effect on

---

[1] See Appendix M, Tables 16 and 17 for the analyses.

[2] See Appendix M, Tables 18 and 19 for the analyses.

TABLE 10

MEANS AND STANDARD DEVIATIONS FOR GROUP SCORES ON PRICE SCALE FACTORS

FACTOR	Videotape Group				Audiotape Group				Hawthorne Control Group			
	Pretest		Post-test		Pretest		Post-test		Pretest		Post-test	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
F_1	6.31	1.53	7.81	1.26	6.31	1.24	7.51	1.26	5.98	1.60	7.55	1.24
F_2	5.91	1.60	6.14	1.60	5.56	1.53	5.90	1.58	5.57	1.42	5.68	1.58
F_3	6.05	1.71	7.22	1.74	6.16	1.47	6.59	1.50	5.68	1.86	6.23	1.83
F_4	6.34	1.63	7.39	1.46	6.38	1.59	6.91	1.71	6.10	1.71	6.66	1.87
F_5	6.20	1.27	6.96	1.26	6.10	1.09	6.58	1.29	5.98	1.34	6.62	1.24
F_6	6.05	1.53	7.12	1.37	5.70	1.46	6.46	1.58	5.73	1.59	6.42	1.53

post-test scores on factor #3, Bodily Action. Planned comparisons did demonstrate the superiority of the videotape group over the audiotape group with a t value of 2.796 ($p < .005$, df = 102).[1] The arithmetic difference favoring the audiotape group over the "Hawthorne control" group was not significant ($p < .10$, df = 102). Bodily Action scores for post-test performances for the entire group of subjects were significantly higher than for pretest performances ($p < .001$, df = 102). Also, a significant interaction was observed on BC (instructor-trial measure) ($p < .01$, df = 1, 102), indicating that without regard to feedback treatment, students taught by the instructor with public address background improved substantially more on Bodily Action than students of the instructor with theater background. This was the only significant BC interaction found in analyses of individual Price factor scores.

On factor #4, Personality, which dealt with the speaker's poise and fluency, the videotape group was again significantly higher on post-test speeches than the audiotape group ($p < .025$, df = 102).[2] As in the case of other factor scores, the arithmetic difference favoring the audiotape group over the control group was not significant ($p < .15$, df = 102). The mean post-test score for all subjects was significantly higher than the mean pretest score ($p < .001$, df = 102).

On Language, factor #5, videotape subjects were significantly better in post-test performances than audiotape subjects ($p < .025$, df = 102).[3] The slight arithmetic difference between audiotape subjects and control subjects did not approach significance. Subject post-test speeches were significantly better on Language than pretest speeches ($p < .001$, $df = 102$).

---

[1] See Appendix M, Tables 20, 21, and 22 for analyses.

[2] See Appendix M, Tables 23 and 24 for analyses.

[3] See Appendix M, Tables 25 and 26 for analyses.

The videotape group demonstrated significantly higher scores on Voice, the sixth factor, in post-test performances than the audiotape group[1] ($p < .001$, $df = 102$). No difference was found between the audiotape group and the control group. The mean Voice score for all subjects was significantly higher on post-test performances than on pretest performances ($p < .001$, $df = 102$).

Figures 1 and 2 summarize pretest and post-test scores for the three treatment groups on each of the six Price scale factors. The most important findings of the three-way analyses of variance and planned comparisons on factor scores and total scores are summarized in Table 11. Significant t values for a priori comparisons and accompanying significance levels are given for each of the two null hypotheses rejected by the majority of the analyses. These results summarize analyses of total scores on speech skill using factor weights provided by two different sources, and analyses of the six individual Price scale factor scores.

Since the analyses summarized in Table 11 suggested that subjects on the whole improved more on some factors than on others, a two-way analysis of variance was computed on subject factor gain scores, with the 108 subjects as one dimension, and the six Price scale factors as the other. In order to make factor scale values comparable, individual

---

[1] See Appendix M, Tables 27 and 28 for the analyses.

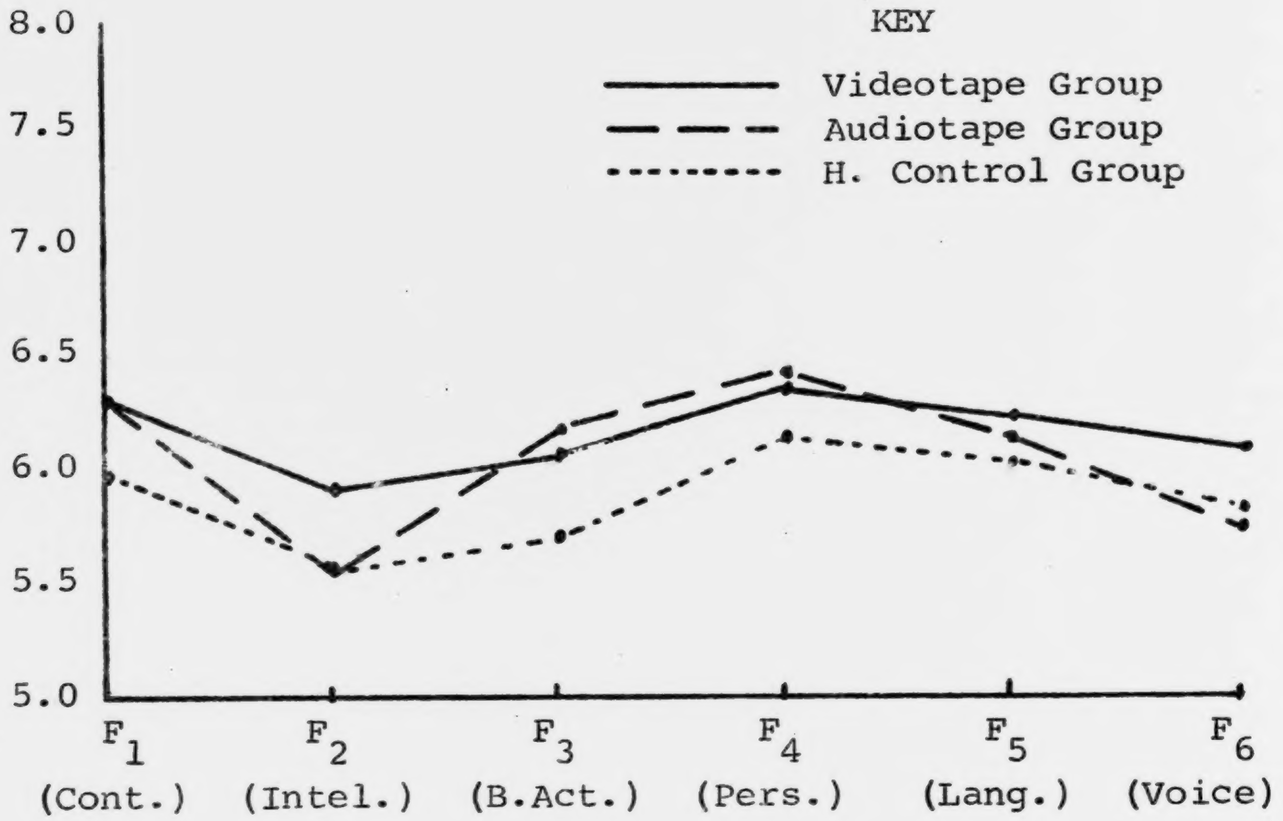

FIGURE 1

PRE-TEST SCORES FOR THREE TREATMENT GROUPS ON PRICE SCALE FACTORS

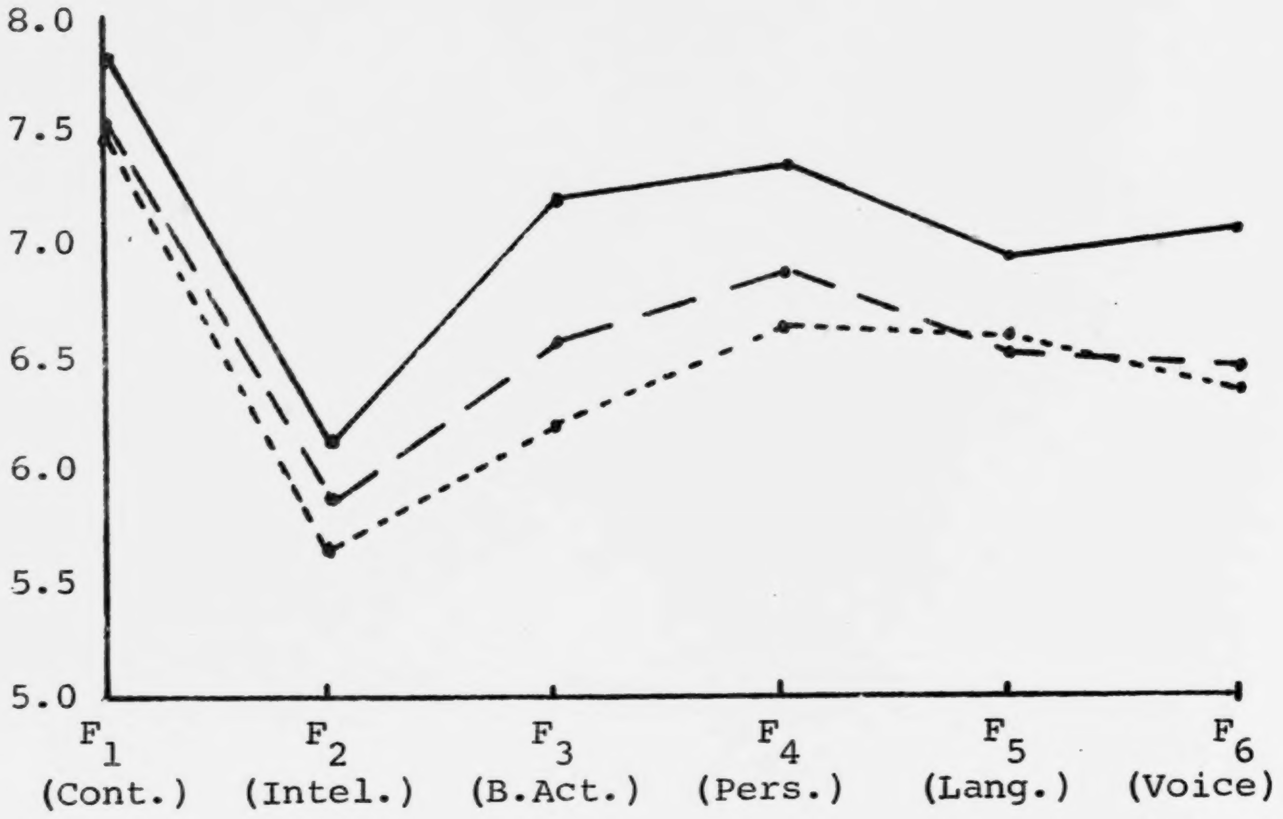

FIGURE 2

POST-TEST SCORES FOR THREE TREATMENT GROUPS ON PRICE SCALE FACTORS

TABLE 11

SUMMARY OF SIGNIFICANT FINDINGS OF 3-WAY$_1$ ANALYSES OF VARIANCE ON FACTOR AND TOTAL WEIGHTED SCORES

Null Hypotheses	$A_1C_2 \leq A_2C_2$ (Vid.post ≤ Aud.post)			$C_2 \leq C_1$ (Post-test ≤ Pretest)		
Scores Analyzed	DF	t	alpha	DF	t	alpha
Total Weighted$_{\text{Price}}$	102	2.570	.01	102	4.332	.001
Total Weighted$_{\text{Judge}}$	102	2.580	.01	102	5.697	.001
#1 Content	102	1.309	NS	102	12.396	.001
#2 Intelligibility	102	1.211	NS	102	1.604	NS
#3 Bodily Action	102	2.796	.005	102	3.196	.001
#4 Personality	102	2.132	.025	102	3.198	.001
#5 Language	102	2.089	.025	102	3.463	.001
#6 Voice	102	3.330	.001	102	4.238	.001

factor gain scores for subjects were divided by the factor standard deviation, then multiplied by the mean factor gain score standard deviation. Table 12 shows the results of this two-way analysis.

TABLE 12

RESULTS OF 2-WAY ANALYSIS OF VARIANCE ON SUBJECT GAINS FOR SIX FACTORS

Source of Variation	DF	SS	MS	F
Subjects	107	606.3175	5.6665	6.992[a]
Factors	5	81.1344	16.2269	20.022[b]
Residual	535	433.6002	0.8105	
Total	647	1121.0521		

[a] ($p < .01$, df = 107,535)

[b] ($p < .01$, df = 5,535)

Following the discovery of significant differences between subject gains on the six Price factors ($p < .01$, df = 5, 535), a post-hoc analysis of these gains was conducted. Results of the Newman-Keuls procedure, chosen for this analysis, are given in Table 13.[1]

TABLE 13

RESULTS OF NEWMAN-KEULS PROCEDURE ON SUBJECT GAIN SCORES FOR EACH FACTOR

Order of Magnitude	1	2	3	4	5	6
Factor Number	2 Intell.	5 Lang.	3 Bod. Act.	4 Person.	6 Voice	1 Content
Mean Gain	0.226	0.624	0.715	0.719	0.844	1.421
	Differences Between Factor Gain Scores					
Factor	2	5	3	4	6	1
Intelligibility: 2	--	0.398^a	0.489^a	0.493^a	0.618^a	1.195^a
Language: 5		--	0.091	0.095	0.220	0.797^a
Bodily Action: 3			--	0.004	0.129	0.706^a
Personality: 4				--	0.125	0.702^a
Voice: 6					--	0.577^a
Content: 1						--
Truncated Range r		2	3	4	5	6
$q_{.99}$ (r, 107)		3.72	4.22	4.53	4.74	4.90
$q_{.99}$ (r, 107)$\frac{MS_e}{n}$		0.322	0.365	0.392	0.410	0.424

$^a(p < .01)$

---

[1] Winer, Statistical Principles, pp. 77-87.

Results of the Newman-Keuls procedure show that during the semester subjects as a group gained significantly more on factor #1, Content, than on any other Price factor. This finding was significant at the .01 level. Also, subjects progressed significantly less (.01) on the second factor, Intelligibility, than on any other factor. No statistically significant differences were found between gains on "Bodily Action," "Personality," "Language," and "Voice." Since this analysis grouped all subjects together without regard to treatment group or instructor, its only meaning is in terms of general improvement by subjects on each factor.

Results of Analyses of Subsidiary Measures

Subsidiary measures of the dependent variable in this study included point totals for subjects' speeches and total course grades, attendance at recitation section meetings, number of conferences with instructor, and number of advanced courses in speech taken after the completion of fundamentals of speech. Compared to the measurement provided by nine judges using the Price scale, these measures were considered relatively imprecise. For this reason they were analyzed using F tests, and Tukey post-hoc comparisons rather than the more powerful planned comparison t test and Newman-Keuls post-hoc procedure.

Subject speech points and course points were analyzed by means of analysis of variance. Since it was questionable whether the effect of the two instructors could be properly compared by analysis of either raw grade scores given by each instructor or scores adjusted by linear transformation, separate one-way analyses of variance were undertaken for each instructor. These investigated the possible treatment effect on the scores subjects received on their five graded speeches and total course points.

Of the four analyses, the only significant difference between treatment groups was on total course points given by instructor #2, the instructor with theater background ($F = 3.8075$, $p < .05$, $df = 2, 51$).[1] A Tukey post-hoc comparison on instructor #2's course points showed that his videotape group was significantly higher in total course points than his audiotape group or control group ($p < .05$, $df = 51$).[2]

Other subsidiary data which were viewed for group differences included class attendance, conferences with instructors, and advanced course enrollment.[3] No significant differences were found for any of the three subsidiary measures for simple effects of treatments or instructors, or on treatment-instructor interactions.

---

[1] See Appendix N, Tables 30-33 for the four one-way analyses of variance.

[2] Winer, Statistical Principles, p. 87.

[3] Although the basic graphic form of these three sets of measures was essentially that of a "J-curve" rather than a normal curve, Norton's study indicates the propriety of analysis of variance under these circumstances: "It is evident . . . that the F-distribution is amazingly insensitive to the form of the distribution of criterion measures in the parent population, granting that the same form is common to all treatment populations. Discrepancies [between F-table significance levels and empirically determined levels] significant at the 5% level are found only for the leptokurtic and rectangular distributions, and even then the absolute discrepancies are quite small. Apparently, the F-distribution is practically unaffected by lack of symmetry, *per se*, in the distributions of criterion measures. . . ." See Lindquist, Design and Analysis, p. 81. Norton's study, "An Empirical Investigation of Some Effects of Non-normality and Heterogeneity on the F-distribution" (unpublished Ph.D. thesis in education, State University of Iowa, 1952), is discussed pp. 78-90.

Finally, the questionnaire given all subjects at the final examination period was tabulated and analyzed for significant patterns. Responses and results of the two-way analysis of variance conducted for each of thirty-one questions are summarized in Appendix K.

Summary

Using the Ebel intra-class method for estimating reliability of multiple ratings of performances, a coefficient of .95 was found for total weighted judgments of 216 speeches pooled across nine judges. The median intra-class reliability for individual factor scores was .91. Test-retest reliability based on two measurements of twelve speeches, judged two weeks apart in two different orders, was .93. These high reliability coefficients, determined by two different methods, were taken as indications that the measurements of the dependent variable were of sufficient precision and stability to warrant their analysis for determining experimental effects.

No attempt was made to compute validity coefficients for the speech ratings. However the high degree of agreement demonstrated by experienced judges, along with Price's work on validating the instrument, suggested that the measurements were valid estimates of speech skill.

The assumptions underlying analysis of variance, including random subject assignment, homogeneity of variance, and normality of group criterion measures, were satisfied in terms of this study. Winer's Case II model for three-way analysis of variance with repeated measures on one variable formed the basis for statistical analysis. Planned comparison t tests and F tests were conducted as part of the a priori analysis.

Results of the Case II analyses of total speech scores for subjects, computed with two different sets of factor weights, showed the videotape group was significantly higher at the .01 level in post-test performance than the audiotape group on that performance. However the audiotape group was not significantly higher than the "Hawthorne control" group on total scores. Combined subjects were significantly better on overall speech skill in post-test performances than in pretest performances, at the .001 level.

Analyses of individual factor scores revealed that in post-test performances, the videotape group was significantly better than the audiotape group on the factors of Bodily Action, Personality, Language, and Voice. Significance levels for these differences were .005, .025, .025, and .001 respectively. The audiotape group was not significantly better in post-test performances than the Hawthorne control group on any of the six Price factors. Subjects, without regard to treatment or instructor, were significantly better in post-test speeches than in pretest at the level of .001 on all factors except Intelligibility. A post-hoc analysis showed that during the semester subjects had gained significantly more on Content than on any other factor, and significantly less on Intelligibility.

Little evidence of direct bearing on the hypothesis was uncovered by analysis of subject speech grades, course grades, attendance, instructor-conferences, advanced speech course registration, or post-test questionnaires.

In summary, results of a priori comparisons show that on the whole the videotape group performed significantly better in post-test speeches than the audiotape group, whose performances were not significantly different from those of the "Hawthorne control" group. Clearly, videotape feedback had a positive impact upon student performances.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

The primary purpose of this experimental study was to answer the question: Does videotape feedback help students achieve greater skill in speaking than they otherwise might acquire in a speech class not employing this technique? To provide the answer, a general hypothesis based on a theoretical model for feedback was established: "The greater the completeness and accuracy of student speech performance feedback, the greater the degree of speech skill a student will later exhibit." For purposes of the investigation, feedback was defined as "any consequences or results of performance that are perceived by the learner."

As the first of the three independent variables, feedback was operationalized in fundamentals of speech classes as three levels of electronic replay: (1) videotape replay of first and fifth round class performances plus conventional criticism on all performances; (2) audiotape replay of first and fifth round class performances plus conventional criticism on all performances; (3) "Hawthorne control" activity of viewing videotape replay of other speakers plus conventional criticism on all performances.

The subjects' two instructors, representing dissimilar interests and backgrounds, represented the second independent variable: (1) public address and group communication; (2) theater. The measurements of two trials for each subject represented the third independent variable: (1) pretest performance; (2) post-test performance. These three levels of feedback, two instructors, and two measurements represented a 3 X 2 X 2 fixed effects model.

Subjects were 108 Eastern Michigan University students enrolled in seven sections of Speech 124, Fundamentals of Speech, during the Spring Semester, 1967. Random assignment to the three feedback conditions was based on four strata: age, sex, speech background, and instructor. Besides making the subject assignment to each treatment group equal on these strata, this procedure resulted in the assignment of an approximately equal number of subjects from each section to each treatment.

To measure the dependent variable, speech skill, five faculty and four student judges were trained in the use of an eleven-point version of the Price speech performance rating scale. Videotaped performances were stratified on the three levels of feedback, two instructors, and two measurements, and randomly ordered in groups of twelve. Judging of the 216 speeches was accomplished in ten rating sessions with an overall reliability of .95. Subsidiary data including speech scores, course points, attendance, advanced speech course registration, and questionnaire responses were also gathered for analysis.

Ratings of the judges were subjected to a Winer Case II three-way analysis of variance with repeated measures on one variable. Planned comparison t tests were employed where the theoretical model predicted particular relationships in the form of research hypotheses.

Conclusions

Analysis of experimental data provided a test of the major hypothesis regarding the impact of electronic feedback on student demonstration of speech skill. This analysis also produced answers to questions inherent in the three-factor design employed in testing the major hypothesis. These questions involve possible differences between subjects taught by the two instructors and between subject pre-test and post-test performances. The question of the relative reliability of ratings by student and faculty judges was also answered.

Of greatest interest were the findings regarding the major experimental hypothesis establishing theoretical relationships between the three levels of electronic feedback. The predictive success of these two hypothesized relationships will be discussed first.

Hypothesis: Students who view videotape recordings of two of their own class performances will demonstrate greater speech skill in post-test speeches than students who hear audiotape recordings of two of their class performances. This hypothesis was clearly supported by statistically significant results in this study. Overall speech ratings showed that subjects who viewed videotapes of their first and fifth class speeches were appreciably better than those who heard audiotape recordings of their first and fifth performances. This finding was based on analysis of variance of the subjects' total weighted speech scores for pretest and post-test performances, significant at the .01 level. Viewed in terms of weighted total speech score gains between pretest and post-test performances, the videotape group improved forty per cent more than the audiotape group. Therefore, the improvement in speech skill brought about by the two twenty-minute exposures to videotape feedback during the semester was both statistically significant and pedagogically meaningful.

In order to provide more information on what dimensions of subject speech skill were most affected by videotape, analyses of scores on each of the six Price scale factors were undertaken. During final speeches, members of the videotape group were significantly better than members of the audiotape group on four of these factors: Bodily Action, Personality, Language, and Voice. However no significant differences were found between the videotape and audiotape groups on two factors: Content and Intelligibility.

These findings suggest that videotape feedback had a greater positive impact on delivery aspects of speech performance than on composition aspects of speech skill. This conclusion is in keeping with results of the microteaching research of Allen and Fortune. The reason for this observed difference in impact may come from the fact that beginning speech students seem to find it easier to focus attention on problems of delivery than on problems of content, and that videotape replay seems to aid such attention. Still it must be remembered that on weighted total Price scale scores in which content played an important role, the videotape group was significantly better in overall speech skill.

This demonstration of the pedagogical superiority of videotape feedback is the most far-reaching result of this study. It is consistent with the observation of many speech teachers and provides backing for the unsupported written claims for the procedure discussed in Chapters I and II. Moreover, it provides a quantitative basis for further research on methods of maximizing the positive impact of videotape replay in fundamentals of speech classes.

Determining the best ways of employing videotape feedback requires an understanding of the reasons for the superiority of this procedure over other methods of feedback. The theoretical model established in Chapter I indicates that the functions of feedback are to provide information and reinforcement for the learning process. Although this study did not attempt to establish reasons for the positive impact of videotape, analyses of scores on individual Price scale factors, records of attendance, instructor conferences, and subject questionnaire responses offer several clues.

The significant difference between videotape subjects and audiotape subjects on factor #3, Bodily Action, might be explained by the fact that videotape feedback provides information about a speaker's movement, gestures, and facial expression while audiotape feedback does not. However this explanation does not provide a rationale for the finding that the videotape group was significantly better on factor #6, Voice, since both groups heard comparable sound tracks. This finding suggests that the greater feedback completeness of videotape provides greater reinforcement of speech skill in general, although the possibilities of "halo effect" in ratings and high correlation between

Voice and Bodily Action may also provide a partial explanation. Although not statistically significant, class attendance figures indicated that members of the videotape group were absent arithmetically fewer times than audiotape subjects. The videotape group also requested more conferences with their instructor than the latter group, although again the difference was not statistically significant. To the extent that these differences are indicative of differential impact of feedback mechanisms, they provide some evidence for the relative importance of the motivational, and therefore reinforcement function of videotape feedback. Further research is needed to expand our knowledge of this phenomenon.

It does not appear that differences found can be explained by the length of time each subject spent in his feedback treatment, since audiotape subjects spent as much or more time in the playback rooms than videotape subjects spent in the television studio. Nor can it be accounted for on the basis of the "Hawthorne effect," the tendency for subjects to increase performance when they feel they are receiving special attention from the researchers. Questionnaire responses indicated that subjects from all three groups apparently felt that they were receiving the same amount of attention and interest from those whom they thought controlled the experiment. However responses also showed that all groups felt that viewing videotapes of performances would be more helpful than hearing audiotapes. Although the reasons for the greater pedagogical value of videotape feedback are still unclear, the existence of that value within the conditions of this study is not.

Hypothesis: Students who hear audiotape recordings of two of their own class performances will demonstrate greater speech skill in post-test speeches than students who view videotape recordings of three speakers other than themselves. Although arithmetic differences supporting this hypothesis were found, no statistically significant evidence was provided by this study for acceptance of this hypothesis. Members of the audiotape and "Hawthorne control" groups displayed speech skill in post-test performances which failed to differ in a statistically significant manner. On the basis of this finding, it must be assumed that any effect hearing audiotape replay of two class speeches had on later performance was equaled by viewing and evaluating performances of the three other speakers. It is also possible that neither treatment had any real effect on performance.

The similarity of speech skill demonstrated by the two groups may have been partly the result of the fact that the two treatments were employed at different times in the semester. Audiotape members heard their first and fifth class speeches, the latter at least one month before their final or "post-test speech." The "Hawthorne control" subjects viewed videotapes of the three speakers other than themselves just one week before they gave their final speeches. Since the effect of an experimental treatment tends to diminish with the passage of time, the closer proximity of the "Hawthorne control" activity to the post-test performance may have caused this group to perform at a higher level than they would have, if they had received their treatment at the same time as the audiotape group. It is impossible to determine to what extent the similarity of audiotape and "Hawthorne control" subjects' post-test performances was the result of the different time of treatment application.

This finding of no significant difference between the final speeches of the audiotape and "Hawthorne control" groups need not necessarily be interpreted as indicating that audiotape replay has no value in teaching speech. Both the theoretical model for feedback established in Chapter I and observations of many speech teachers suggest that audiotape is a valuable teaching aid. In this study, the audiotape group was arithmetically higher on overall speech skill in post-test performances than the "Hawthorne control" group. Also, subject questionnaire responses showed that audiotape members felt their experimental activity was significantly more helpful than did members of the "Hawthorne control" group. As suggested above, the confounding of the treatment with the time of application of treatment makes it impossible to determine whether the hypothesized value of the technique exists in reality. It is also possible that both hearing one's own performances on audiotape and seeing those of others on videotape are both useful teaching aids.

On the other hand it must be noted that this finding is consistent with the only other experimental study testing the pedagogical value of audiotape feedback in a begining speech class. As shown in Chapter II, Nystrom and Leaf found no significant differences between the experimental group which heard sound recordings of their speaking and the control group which heard no recordings. In the present study, no significant difference was found between the group which heard sound recordings of their speeches and the group which viewed videotape recordings of other speakers. Further research is needed to resolve questions about the usefulness of audiotape replay in teaching speech.

Hypothesis: Students will demonstrate greater speech skill on post-test speeches than on pretest speeches, without regard to feedback treatment or instructor. This hypothesis was clearly substantiated on the basis of data from this study. Subjects from all treatment groups were significantly better at the .001 level on total weighted scores for post-test speeches than for pretest performances. Therefore, although the videotape group was significantly better than audiotape and control groups on their post-test scores, all three groups improved significantly in overall speech skill.

Analysis of subject gain scores on individual Price scale factors indicated significant improvement for all groups on five of the six factors: Content, Bodily Action, Personality, Language and Voice, with greatest improvement on Content. No statistically significant improvement was shown on the factor of Intelligibility. This finding indicates that the overall impact of this particular fundamentals of speech course caused subjects to make their greatest improvement on Content, less improvement on Bodily Action, Personality, Language, and Voice, and no significant improvement on Intelligibility.

Analysis of subject gains on total scores, without regard to instructor or treatment, clearly showed that some subjects improved significantly more than others, at the .01 level. Even though subjects as a whole showed significant improvement on overall speech skill, negative gains were found for some individuals, although these could have been the result of subject variability of performance and measurement error.

Question: Are there differences between post-test speech ratings of subjects taught by the instructor with public address background and those taught by the instructor with theater background? No statistically significant differences were found between final performances of students of the two instructors, either on total weighted scores, or on individual Price scale factors.

A slight interaction effect between instructor and level of feedback was noted on total speech scores. This suggests that one type of instructor may be more effective employing one type of feedback procedure, while another may find another procedure more useful. This result is diluted by the fact that a nearly significant interaction effect between treatment and instructor occurred by chance on pretest performances before the impact of treatment could have been felt. To correct for this sampling bias, subject gain scores were inspected, showing that any true interaction effect probably existed between the audiotape and control groups and the two instructors. The instructor with background in public address appeared relatively more effective employing the audiotape feedback method, while the instructor with background in theater appeared relatively more effective with the videotape replay of other speakers. Since mean gain scores were identical for both instructors' videotape subjects, this feedback procedure may have had a stabilizing effect on the instructor's impact. These observations, however, can be stated only as a tentative implication of the data, requiring further study.

A clearly significant interaction was seen at the .01 level between instructor and pre-post measure on factor #3, Bodily Action. Although students of both instructors achieved the same Bodily Action mean scores on pretest performances, the public address instructor's students were clearly superior to the theater instructor's students on that factor on post-test performances. However, since only one instructor from each background was tested, this statistically significant finding can not be freely generalized to other individual fundamentals of speech instructors coming from these two backgrounds.

Question: Can specially trained advanced speech students judge speech performances with as high reliability as specially trained speech faculty members? Results clearly demonstrated that highly motivated, specially trained students can make judgments which are as reliable as those made by speech faculty members receiving the same training. This finding suggests that advanced speech students can effectively be used to evaluate performances in speech research. It also offers the possibility that for the sake of economy such students might be used within the fundamentals of speech instructional program, although this contingency requires substantial additional research.

Future Research

Results of this study suggest a number of questions which provide the bases for future research. The majority of these questions apply to ways of maximizing the positive effects of videotape feedback for teaching speech:

1. What is the process by which videotape feedback helps a student improve his speech skill? A series of experiments should be conducted to establish a theoretical model explaining how this procedure works within the speech instructional process.

2. What characteristics of the student determine

whether videotape feedback will be pedagogically helpful, neutral, or harmful? The significant difference among individual subject gains, computed without regard to treatment, suggests the possibility that like any other teaching method videotape replay may be without value or even harmful to some students. Of course a great many other variables such as subject variability and measurement error could account for this difference in gains measured for different subjects. Still it would be valuable to know how to identify those students for whom the technique is most helpful, and those for whom it is least helpful.

3. At what point in the semester is the use of videotape feedback maximally effective for inducing desired improvement in performance? This should be investigated both for short term and long term gains.

4. What is the best number of videotape feedback experiences for most three-hour fundamentals of speech courses? Performance curves should be established to determine how many exposures are needed to reach the point of diminishing returns. The expense of this procedure requires that the amount of improvement for each use be known.

5. Is it better to give students their instructor's critiques of their performances before they have viewed their videotapes or after? Although previous research has provided some evidence on this question, the definitive answer has yet to be established.

6. Does the student benefit from viewing a classmate's speech before his own? If so, what criteria should be employed in choosing that speech? It appears likely that playing a classmate's speech does more than merely "validate" the videotape medium for the student, but this

process is not fully understood at the present time.

7. Would the videotape feedback procedure be equally valuable for advanced public speaking classes? Inherent in this question is the possibility that the technique may be more valuable or less valuable. By introducing other subject variables such as speech training and experience, research designed to answer this question might shed further light on the process by which this procedure works to improve speech skill.

8. Is the amount of improvement brought about by videotape feedback worth the high cost of the procedure? This is a difficult administrative problem for which the results of this study, statistically significant as they are, fail to provide an answer.

9. Can the videotape replay technique be employed by trained advanced speech students in conjunction with an instructor in order to teach effectively more students with fewer faculty members? With the growing shortage of teachers, this question has broad national significance.

10. Is audiotape feedback more valuable than no electronic feedback? Because of the relative ease and lack of expense of employing audiotape replay in the teaching of speech, this question deserves attention. The study reported here fails to provide a definitive answer.

11. Can the Price speech performance rating scale be further refined, and the process for using it be improved? With so much speech research employing "speech skill" as the dependent variable, it is imperative that this instrument be made as reliable and valid as possible. Work must also continue on designing and testing other, more precise instruments for research tools.

12. Finally, will a replication of this study lead to similar results? In order to increase our ability to generalize from the findings of this study, it would be valuable for that replication to involve students and instructors representing backgrounds different from those employed in this project. If the number of available subjects is sufficiently high, it would be useful to include a fourth treatment in that replication, a pure control group who engages in no experimental activity other than writing self-evaluations.

Although this study has demonstrated that fundamentals of speech students who view videotape recordings of their class speeches demonstrate significantly greater speech skill at the end of the semester, further research is needed to provide greater understanding of the process and methods for maximizing this positive impact upon student performance.

APPENDICES

APPENDIX A*

SPEECH PERFORMANCE RATING SCALE

SPEAKER NUMBER:____________________ JUDGE:__________________

Rate the speaker on each of the six scales by CIRCLING the number you think is appropriate. DO NOT MARK BETWEEN THE NUMBERS.

1. IS THE CONTENT OF THE SPEECH CLEAR AND REASONABLE?
   (reasoning, supporting material, concreteness, clearness of purpose, use of motive appeals, appropriateness of topic, organization)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

2. IS THE SPEAKER INTELLIGIBLE?
   (articulation, pronunciation)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

3. DOES THE SPEAKER COMMUNICATE WELL THROUGH BODILY ACTION?
   (movement, gesture, facial expression, eye contact)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

4. WHAT IS YOUR IMPRESSION OF THE SPEAKER'S PERSONALITY?
   (personal appearance, posture, poise, fluency)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

5. DOES THE SPEAKER USE LANGUAGE VIVIDLY AND IMAGINATIVELY?
   (use of figurative language, word choice, originality of ideas)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

6. DOES THE SPEAKER HAVE A PLEASING AND EXPRESSIVE VOICE?
   (pitch, vocal quality, rate)

1	2	3	4	5	6	7	8	9	10	11
poor					average					good

---

*Eleven step version of Price scale. Employed to measure this study's dependent variable.

APPENDIX B*

VIDEO TAPE REPLAY ACTIVITY

Self-Evaluation Form

I. Before the next class recitation period, please do the following:

  A. Go to the Roosevelt TV studio (Room #108) at one of the two times specified below:

    1. Two times available for viewing:

      a. ______________________________

      b. ______________________________

    2. Bring pencil and paper for note taking.

    3. Sign the special sign-up sheet.

    4. You will see a classmate's speech once, and your own speech twice.

      a. Take notes on your speech.

    5. Be sure your name has been recorded for attendance.

    6. This activity should take roughly one half hour.

  B. After hearing the tape replay of your speech, write a 1-2 page "Self Evaluation" of your speech performance:

    1. Base this evaluation on all the information you have about your performance. (i.e. class comments, instructor criticism, tape replay of the speech, and your own feelings.)

    2. Use the following format for your self evaluation:

      a. Name:______________ Date of speech:_________

      b. Strongest aspect of the speech performance:

      c. Weakest aspect: (i.e. what needs most work)

---

*Instructions for subjects of pilot study on video-tape recording-replay procedure.

d. Delivery: (You may well want to discuss some of the following)

1. Directness:

2. Vitality:

3. Voice:

4. Bodily action:

e. Composition: (Discuss whatever seems noteworthy)

1. Subject and speech purpose:

2. Introduction:

3. Conclusion:

4. Organization:

5. Supporting material:

6. Language:

3. Be realistic and honest in your self evaluation!

4. Date due: Next recitation period, ____________ at the beginning of class.

C. Your instructor will hand back your speech outline with his evaluation of your speech, only after you hand in your self evaluation.

APPENDIX C*

124 TV FUNDAMENTALS OF SPEECH

Questionnaire: Use of Video Tape for Self-Analysis

Procedure:

1. Answer each question according to your honest belief or feeling.
2. Do not sign this questionnaire.
3. Circle the response on the 7-point continuum which best reflects your belief regarding each question.

e.g.

Definitely Yes	Yes	[✓ Probably Yes]	Undecided	Probably No	No	Definitely No

4. Your answers will help improve the teaching techniques in the fundamentals of speech courses in semesters to come.

Questions:

1. Did you know that you were being video taped?

	Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No
[Responses:	5	3	1	0	0	1	4]

---

*Questionnaire text and responses for pilot test of videotape recording-replay procedure.

2. If you did know that you were being video taped, did that knowledge make you more nervous? (Mark only if applicable.)

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	1	1	1	1	3	3]

3. If you did not know you were being video taped do you think that such knowledge would have made you more nervous during the video taping? (Mark only if applicable.)

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	1	4	0	1	2	1]

4. Including watching your own performances and writing your self evaluation did you spend over one hour in the total activity?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	6	1	1	1	1	2]

5. Do you think that this use of video tape was worth your time spent?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	6	5	2	0	0	0	0]

6. Did you see your performance twice and that of another student's once?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	6	2	0	0	0	1	4]

7. Do you think that it would be helpful to see your performance more than twice?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	2	1	2	1	3	2	2]

8. Was video tape valuable for delivery?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	5	4	2	3	0	0	0]

9. Was it valuable for composition?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	3	4	4	2	0	0]

10. Was it helpful to see the performance of another classmate?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	5	2	1	1	4	0]

11. Would it have been just as valuable if you had not been asked to write a written self critique?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	4	1	1	5	2	0]

12. Would it have been helpful if we had taped your first speech for you to see and hear?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	5	4	0	2	1	2	0]

13. Would you like to be able to practice a speech and have it video taped for you to see before giving in a class?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses: 5	3	3	0	0	1	1]

14. Did viewing the tape in Roosevelt 108 prove distracting?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses: 1	0	1	1	2	5	3]

15. Would it have been helpful to have the instructor's written critique on your speech while you watched it?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses: 4	3	4	1	1	1	0]

16. Did this experience make you more aware of your problems in oral communication?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses: 7	6	0	1	0	0	0]

17. After seeing yourself do you find that you disagree with some of the criticisms made by the instructor?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses: 0	3	0	1	1	5	4]

18. Did writing a self evaluation make the video tape experience more meaningful?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	7	2	0	0	2	1]

19. On the basis of it do you think that you will actually do anything different in preparing for your final speech?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	3	3	4	2	1	1	0]

20. Do you think that it was unfair for your instructor to require you to see the tape?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	0	0	1	0	6	7]

21. Was the engineer at the viewing studio in Roosevelt helpful?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	8	0	1	2	1	0]

22. Do you think audio taping (just sound) would be as valuable as video taping?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	0	0	1	2	4	7]

23. Do you think that the TV picture and sound gave a fair representation of how you looked and sounded?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	4	4	3	0	0	0	1]

24. Do you think that video taping of your first speech would have made any difference in how much and how you prepared for later speeches during the semester?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	1	5	4	1	0	2	0]

25. If this speech recording had been done in the classroom do you think that the presence of the camera would have been distracting?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	3	1	4	2	3	0]

26. Did basic principles of speech seem to apply to this speech via television?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	3	8	1	1	0	0	0]

27. Do you think that video taping is generally a good technique in a fundamentals of speech course?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	6	7	1	0	0	0	0]

28. Did you have a higher opinion of your performance before you saw the tape than you had after seeing it?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	5	2	1	2	1	2	1]

29. Do you think that it was unfair for the instructor to require you to write a self evaluation?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	0	0	1	0	3	6	3]

30. If you had not been required to see this video tape do you think that you would have done so anyway?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Responses:	5	2	3	3	0	0	0]

APPENDIX D*

QUESTIONNAIRE: "PILOT STUDY"

Circle the group to which you belong:

EMU Faculty (instructor or above)	Grad Student	Undergrad Student

Based on your experiences in tonight's pilot study, please answer the following questions. Circle the response most in keeping with your honest feelings regarding each.

Questions:

1. Did you feel that the University of Wisconsin Speech Performance Scale provided a meaningful measure of "speech skill"?

Responses	Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No
[Faculty:	--	1	5	--	2	--	--]
[Student:	1	6	7	3	--	--	--]

2. Did the individual factors (e.g. content, personality, voice, etc.) seem to be reasonable components of "speech skill"?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	3	2	2	1	--	--]
[Student:	1	5	4	1	4	--	--]

3. For each factor, was it possible for you to make seven discriminations? (i.e. 7 point scale reasonable?)

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	3	3	2	--	--	--	--]
[Student:	3	4	--	--	--	1	--]

---

*Questionnaire text and responses for pilot test of Price scale judging procedure.

4. Was it possible for you to make eleven discriminations per factor? (i.e. 11 point scale reasonable?)

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Responses							
[Faculty:	--	3	1	1	3	--	--]
[Student:	--	7	3	3	--	4	--]

5. Which of the two scale lengths, 7 point or 11 point, permitted you to make more precise judgments? 7 point or 11 point.

Responses		
[Faculty:	2	6]
[Student:	11	6]

6. Was the training session long enough for you to become familiar with the 6-factor rating scale?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Responses							
[Faculty:	3	4	--	--	1	--	--]
[Student:	10	5	--	--	1	--	--]

7. Was the training session helpful in increasing your ability to use the scale with precision?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Responses							
[Faculty:	3	1	3	--	1	--	--]
[Student:	5	8	1	3	--	--	--]

8. Did the lecture-discussion method used in the training session seem the appropriate method for training judges?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Responses							
[Faculty:	3	2	2	1	--	--	--]
[Student:	3	4	6	2	--	1	1]

9. Was the technical quality of the video tape recordings so poor that it seemed to hinder the precision of your judgments?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	--	--	--	1	1	4]
[Student:	--	--	--	--	3	10	6]

10. To what extent was the picture and sound quality distracting?

Responses	Made Judging Impossible	Quite Distracting (Judging Possible)	Mildly Distracting	Not Distracting
[Faculty:	--	--	3	5]
[Student:	--	--	7	10]

11. Were you getting so tired and/or bored toward the end that your evaluations of the last speeches were less precise than earlier evaluations?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	1	--	5	--	1	1	--]
[Student:	1	3	2	2	4	4	1]

12. Do you think that this scale would be useful in speech research?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	2	4	1	1	--	--]
[Student:	2	7	4	3	1	--	--]

13. Do you feel that this scale would be useful in classroom teaching?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	--	--	2	4	2	--]
[Student:	2	9	2	2	--	--	2]

14. Would training in the use of this scale help inexperienced teachers understand what makes good speech performances?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	1	5	1	1	--	--]
[Student:	2	7	4	2	1	1	--]

15. Do you think this 6-factor scale is a better measure of speech skill than the single-factor scale of "General Effectiveness"?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	3	--	2	1	--	1	1]
[Student:	3	7	6	--	--	1	--]

16. Does this scale make for more precise evaluation of overt speech skill than the method used in the speech classes you teach or take?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	1	--	--	--	4	3]
[Student:	1	6	3	4	1	1	1]

17. What parts of the scale, if any, did you find confusing?

[Responses on this and the following "open ended" questions are too diverse to be summarized here.]

18. Do you have any suggestions for improving the training session?

19. Do you have any suggestions for improving the use of this scale in speech research?

Thank you for your fine cooperation in this research project!

John Sattler

APPENDIX E*

Department of Speech & Dramatic Arts
February 3, 1967

From: John W. Sattler, Chairman

This semester, the Department of Speech and Dramatic Arts is cooperating with the Office of Instructional Broadcasting of EMU in a research project. This project is sponsored by the United States government. Its findings could have nation-wide significance. It involves the attitudes of fundamentals of speech students toward selected audio-visual techniques.

Below is a list of the Fundamentals of Speech sections involved in this research. All of you will have a part in this project, in one way or another. Each of you will be given written instructions several times this semester, and it is essential that you carry them out exactly as specified. At the end of the semester, you will be asked your opinions regarding these techniques.

On behalf of the Department of Speech and Dramatic Arts, may I say that it is our hope that those involved will give their full cooperation to this governmentally sponsored project.

Thank you for your cooperation,

John W. Sattler, Chairman

[Sections involved: Speech 124--1, 3, 4, 5, 6, 7, and 8.]

---

*Text of general instructions and rationale read to all subjects at the first class meeting.

APPENDIX F*

DEPARTMENT OF SPEECH AND DRAMATIC ARTS

Questionnaire for Fundamentals of Speech Students: 121 & 124

Name: ______________________________Course:______Section:____

Instructions

It is the purpose of this questionnaire to provide information for the speech department about the background of each fundamentals of speech class. Your answers can help improve the course you are now enrolled in this semester.

1. Put an "X" next to the answer you wish to give for each question.
2. Answer all questions, as accurately as possible.

Questions

1. What class or year in school are you?

   1st semester Frosh:_______ 2nd sem. Frosh.________

   1st sem. Soph:______ 2nd sem. Soph:____ Jr:____ Sr:____

2. How old are you?

   17:____ 18:____ 19:____ 20:____ 21 or older:____

3. What career do you plan to follow?

   Business:___ Teacher:___ Law or Medicine:___ Clerical:___

   Sales:___ Housewife:___ Undecided:___

4. Have you ever traveled outside the U.S.A., other than to Canada?

   Yes:_____ No:_____

---

*Text of pretest questionnaire for experimental subjects, given first day of class.

5. Are you married?

Yes:_____ No:_____

6. Do you think that the way a person speaks has any effect on how good a job he can get?

Yes:_____ Undecided:_____ No:_____

7. Do you think this fundamentals of speech course will be harder for you than for most other students now enrolled in it?

Definitely Yes:____ Probably Yes:____ Undecided:___ Probably No:____ Definitely No:___

8. Have you had a high school course which devoted a semester or more to speech? (e.g. giving speeches, acting, radio, etc.)

Yes:_____ No:_____

9. Do you think that reading a speech textbook can help you learn to speak more effectively?

Definitely Yes:____ Probably Yes:___ Undecided:___ Probably No:___ Definitely No:___

10. Do you think that speakers look and sound the same over television as they do when you see them speaking in person?

Definitely Yes:____ Probably Yes:___ Undecided:___ Probably No:___ Definitely No:___

11. Have you ever held an office which required you to speak before groups of fifteen people or more? (e.g. school clubs, etc.)

Yes:_____ No:_____

12. What is your best guess as to the grade you will receive in this speech course?

A:____ B:____ C:____ D:____ E:____

13. Were you ever in a school or church play?

Yes:_____ No:_____

14. Did you have a high school public speaking course?

Yes:_____ No:_____

15. Do you think the course you are now enrolled in (121 or 124) will be harder for you to get a good grade in than your other courses this semester?

Definitely Yes:____ Probably Yes:___ Undecided:___ Probably No:___ Definitely No:___

16. What type or types of speech courses did you take in high school? (i.e. solely devoted to speech.) Check all categories that apply:

Acting:___Public Speaking:___Radio:___Debate:___None:__

17. Does the thought of giving speeches in front of an audience make you feel uncomfortable?

Definitely Yes:___ Probably Yes:___ Undecided:___ Probably No:___ Definitely No:___

18. Have you ever appeared on radio or television?

Yes:_____ No:_____

19. Did you engage in forensic contests in high school? (e.g. Oratory, Declamation, Dramatic Reading, Debate, etc.)

Yes:_____ No:_____

20. Do you think there is much difference of opinion among speech experts as to what constitutes a "good speaker"?

Definitely Yes:___ Probably Yes:___ Undecided:___ Probably No:___ Definitely No:___

THANK YOU FOR YOUR COOPERATION.

APPENDIX G*

FUNDAMENTALS OF SPEECH

Name:____________________ Sect.:___ Date of Speech:___________

I. Before the next class recitation period, please do the following:**

  A. Go to the TV studio in Roosevelt Hall, Rm. 108, at the following time and day:______________________

    1. If you are not able to go at the appointment time, go now to Roosevelt Rm. 108 and sign up for another time.
    2. Bring pencil and paper for note taking.
    3. Come alone.
    4. Sign the special attendance-tape sign-out card.
    5. You will receive further instructions there.
    6. This activity should take roughly one-half hour.
    7. Be sure you have signed the attendance card!

  B. After seeing the tape replay of your speech--write a 1-2 page "Self-Evaluation" of your speech performance:

    1. Base this evaluation on all the information you have about your performance. (i.e. class comments, instructor criticism, tape replay of the speech, and your own feelings.)
    2. Use the following format for your self-evaluation.

      a. Name:____________ Sect:___ Date of Speech:___
      b. Strongest aspect of the speech performance:

---

*Three different sets of instructions for the three experimental treatment groups. After they had spoken each subject picked up the instruction sheet with his name printed at the top.

**These instructions were given members of Group I.

c. Weakest aspect: (i.e. what needs most work)

d. Delivery: (You may well want to discuss some of the following:

    1. Directness:

    2. Vitality:

    3. Voice:

    4. Bodily action:

e. Composition: (Discuss whatever seems noteworthy)

    1. Subject and speech purpose:

    2. Introduction:

    3. Conclusion:

    4. Organization:

    5. Supporting material:

    6. Language:

3. Be realistic and honest in your self-evaluation!

4. Date due: Next recitation period, ___________, at the beginning of class.

C. Your instructor will hand back your speech outline with his evaluation of your speech, only after you hand in your self-evaluation.

FUNDAMENTALS OF SPEECH

Name:________________ Sect.:____ Date of Speech:___________

I. Before the next class recitation period, please do the following:

   A. Go to the Audio-Visual Center in the basement of the new library.*

      1. Hours:

         a. Monday-Thursday: 7:45 a.m. to 8:30 p.m.

         b. Friday: 7:45 a.m. to 5:00 p.m.

         c. Weekend: Closed.

      2. Bring pencil and paper for note taking.

      3. Come alone.

      4. Sign the special attendance-tape sign-out card at the desk.

      5. Take the two tapes (yours and a classmate's) to one of the two "Study Rooms for the Blind" in the basement. The tape recorders are there.

         a. Listen to your classmate's tape once.

         b. Then, listen to your speech twice.

         c. Take notes on your speech.

         } Listen Alone!

      6. Return the two tapes to the Audio-Visual desk.

      7. Be sure your name has been recorded for attendance.

      8. This activity should take roughly one-half hour.

   B. After hearing the tape replay of your speech, write a 1-2 page "Self-Evaluation" of your speech performance:

      1. Base this evaluation on all the information you have about your performance. (i.e. class comments, instructor criticism, tape replay of the speech, and your own feelings.)

---

*These instructions were given to members of Group II.

2. Use the following format for your self-evaluation:
    a. Name:______________ Sect:__ Date of speech:___
    b. Strongest aspect of the speech performance:
    c. Weakest aspect: (i.e. what needs most work)
    d. Delivery: (You may well want to discuss some of the following)
        1. Directness:
        2. Vitality:
        3. Voice:
        4. Bodily action:
    e. Composition: (Discuss whatever seems noteworthy)
        1. Subject and speech purpose:
        2. Introduction:
        3. Conclusion:
        4. Organization:
        5. Supporting material:
        6. Language:
3. Be realistic and honest in your self-evaluation!
4. Date due: Next recitation period, _____________ at the beginning of class.

C. Your instructor will hand back your speech outline with his evaluation of your speech, only after you hand in your self-evaluation.

FUNDAMENTALS OF SPEECH

Name:________________ Sect.:____ Date of Speech:__________

I. Before the next class recitation period, please do the following:*

  A. Write a 1-2 page "Self-Evaluation" of your speech performance:

    1. Base this evaluation on all the information you have about your performance. (i.e. class comments, instructor criticism, and your own feelings.)

    2. Use the following format for your self-evaluation:

      a. Name:____________ Sect:___ Date of speech:_____

      b. Strongest aspect of the speech performance:

      c. Weakest aspect: (i.e. what needs most work)

      d. Delivery: (You may well want to discuss some of the following)

        1. Directness:

        2. Vitality:

        3. Voice:

        4. Bodily action:

      e. Composition: (Discuss whatever seems noteworthy)

        1. Subject and speech purpose:

        2. Introduction:

        3. Conclusion:

        4. Organization:

        5. Supporting material:

        6. Language:

    3. Be realistic and honest in your self-evaluation!

    4. Date due: Next recitation period, ____________

---

*These instructions were given to members of Group III.

at the beginning of class.

B. Your instructor will hand back your speech outline with his evaluation of your speech, only after you hand in your self-evaluation.

APPENDIX H*

124 FUNDAMENTALS OF SPEECH

"Special Speech Improvement Activity"

____________________________________ Section:____________

You have been selected to participate in a special activity next Monday, May 15th, at 10:00 a.m. The purpose of this 50-minute meeting is to help you improve your speaking. It is open only to those Speech 124 students who receive this instruction sheet with their names at the top.

At 10:00 a.m. next Monday, go to Roosevelt Hall, Room 108 (TV Studio). (There is no lecture that day.) Your instructor will be there to give you full instructions. You need not bring anything to that meeting.

It is important that you remember to attend this activity!

WHAT:	Special Speech Improvement Activity
WHERE:	Roosevelt Hall, Room 108 (TV Studio)
WHEN:	Next Monday, May 15th, at 10:00 a.m.
WHY:	Because it will help you in speaking
HOW:	You'll see!

---

*Instructions to members of Group III to attend the meeting for the "Hawthorne control" activity.

APPENDIX I*

124 FUNDAMENTALS OF SPEECH

Special Speech Improvement Activity

Your Name:________________________________ Section:_____

PLEASE READ THE FOLLOWING:

As indicated earlier, you have been selected to participate in this special activity designed to help you improve your speaking. Today marks the first time this technique will be used at Eastern Michigan University.

I. Description of Theory: The technique used in today's session is based upon research reported in the latest issue of Speech Monographs, Vol. XXXIV, No. 1, March, 1967, published by the Speech Association of America. That research demonstrated that a beginning speech student can improve substantially by viewing a variety of video taped speeches, evaluating those speeches in terms of stated criteria, and comparing each speaker's strengths and weaknesses with his own. It was pointed out that both the evaluation and the comparison must be carried out while the speech is being viewed in order for the student to improve significantly.

II. Description of Today's Activity: You will view video tape recordings of three varied speeches. Each was made before a live audience, so you can judge them accordingly. They are as follows:
   A. A speech 124 student giving a classroom speech.
   B. EMU's President Sponberg, in a portion of his inaugural address to the faculty and students.
   C. Hal Holbrook, playing the part of Mark Twain as he lectured at the turn of the century.

   You are asked to evaluate each speech and compare it to your own speaking.

---

*Text of instructions and rating forms used in "Hawthorne control" activity.

III. Directions for Today's Activity:

A. Write your name and section number at the top of this page.

B. As you view each speech:

1. Evaluate it in terms of the critique sheet provided for each speech. Be as thorough as possible.
2. Think of your own speaking in terms of each item on the evaluation sheet.

C. This activity will last approximately 40 minutes. Please leave your evaluations of the speakers at the front of the room as you leave.

Evaluation Sheet: "Special Speech Improvement Activity"

Circle the speaker evaluated: 1. Student 2. Sponberg 3. Twain

State briefly the speaker's strongest aspect of performance:

State briefly the speaker's weakest aspect of performance:

Discuss whichever of the following criteria seem noteworthy:

I. Delivery:
  A. Directness:
  B. Vitality:
  C. Voice:
  D. Bodily Action:
II. Composition:
  A. Subject and speech purpose:
  B. Introduction:
  C. Conclusion:
  D. Organization:
  E. Supporting material:
  F. Language:

REMEMBER: THINK OF YOUR OWN SPEAKING IN TERMS OF EACH ASPECT OF THE EVALUATION SHEET AS YOU CRITIQUE THE SPEAKER.

Evaluation Sheet: "Special Speech Improvement Activity"

Circle the speaker evaluated: 1. Student 2. Sponberg 3. Twain

State briefly the speaker's strongest aspect of performance:

State briefly the speaker's weakest aspect of performance:

Discuss whichever of the following criteria seem noteworthy:

I. Delivery:

  A. Directness:

  B. Vitality:

  C. Voice:

  D. Bodily Action:

II. Composition:

  A. Subject and speech purpose:

  B. Introduction:

  C. Conclusion:

  D. Organization:

  E. Supporting material:

  F. Language:

REMEMBER: THINK OF YOUR OWN SPEAKING IN TERMS OF EACH ASPECT OF THE EVALUATION SHEET AS YOU CRITIQUE THE SPEAKER.

Evaluation Sheet: "Special Speech Improvement Activity"

Circle the speaker evaluated: 1. Student 2. Sponberg 3. Twain

State briefly the speaker's strongest aspect of performance:

State briefly the speaker's weakest aspect of performance:

Discuss whichever of the following criteria seem noteworthy:

I. Delivery:
  A. Directness:
  B. Vitality:
  C. Voice:
  D. Bodily Action:

II. Composition:
  A. Subject and speech purpose:
  B. Introduction:
  C. Conclusion:
  D. Organization:
  E. Supporting material:

  F. Language:

REMEMBER: THINK OF YOUR OWN SPEAKING IN TERMS OF EACH ASPECT OF THE EVALUATION SHEET AS YOU CRITIQUE THE SPEAKER.

APPENDIX K*

124 FUNDAMENTALS OF SPEECH

Final Questionnaire

Name:______________________________ Recitation Sect. #______

Purpose of Questionnaire:

1. To give 124 instructors your honest reactions to certain aspects of the course. This will help them improve it.
2. To give researchers your reactions to various aspects of the research program begun this semester in 124.

Guarantee to you: For this questionnaire to be valuable, it is necessary that you answer each question honestly and sincerely. Therefore, we make the following promises:

1. Instructional Broadcasting will keep these questionnaires until Monday, June 12th. (Grades must be in by Saturday, the 10th.)
2. No speech 124 instructors, or members of the Department of Speech will be permitted to see any questionnaires until after all course GRADES have been turned in.
3. Therefore, nothing you say here can have any effect on your grade. So please be frank!

---

*Test and responses to post-test questionnaire administered to all subjects at the final examination period. (Only those questions of special interest to this study are tabulated here.)

Format of Questionnaire: There are three sections:

1. Questions dealing with the course. Everyone do these.
2. Questions on specific activities of the research study. There are four parts. Do only the one part which applies to you.
3. General questions on the study. Everyone do these.

Points toward Final Exam: As a bonus for conscientiously filling this out, 15 points will be added to your exam score. (Instructional Broadcasting will send a list to 124 instructors on Friday.)

Answering Questions: Circle the response closest to your belief regarding each question.

Section I. Everyone Answer These

1. Based on your experiences in this course, do you think that lectures via television convey as much material as live lectures?

Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No

2. Does the lecturer (Mr. Mulac) look and sound essentially the same on TV as he does "in person"?

Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No

3. Has the text book helped you in speaking?

Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No

4. Has your instructor's grading of your speeches been generally fair and accurate?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	13.9%	41.7%	28.7%	13.9%	1.9%	--	--

Group Mean Response:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	2.55556
		2 (Audiotape)	2.33333
		3 (H. control)	2.55556
	2 (Instructor)	1 (Instructor #1)	2.74074
		2 (Instructor #2)	2.22222
Grand Mean			2.48148

Two-Way Anovar: $F_B = 8.207$ ($p < .01$, $df = 1,102$)
Instructor #1 $>$ Instructor #2

5. Do you think your instructor has seemed to expect too much of your speeches?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

6. Have the dittoed assignment sheets been generally clear?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

7. Circle any of the assigned speeches you did not think valuable:

#1	#2	#3	#4	#5	#7	#8
Contemporary Problems	Delivery Workout	To Inform	To Persuade	Debate	TV	Contemporary Problems

8. Do you think it is better to have fundamentals of speech a three-hour course, than to have it two-hours?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

9. Do you think that speakers look and sound the same over television as they do when you see them speaking in person?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	0.9%	26.9%	18.5%	11.1%	11.1%	24.1%	7.4%

	Variables	Categories	Means
Group Mean Response:	1 (Treatment	1 (Videotape)	3.97222
		2 (Audiotape)	3.88889
		3 (H. control)	4.33333
	2 (Instructor)	1 (Instructor #1)	4.07407
		2 (Instructor #2)	4.05556
Grand Mean:			4.06481

Two-Way Anovar: Not significant.]

10. Would viewing the video tape of your performance be more helpful than discussing it in private with your instructor?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	20.4%	27.8%	22.2%	9.3%	7.4%	12.0%	.9%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	3.02778
		2 (Audiotape)	3.13889
		3 (H. control)	2.69444
	2 (Instructor)	1 (Instructor #1)	3.05556
		2 (Instructor #2)	2.85185
Grand Mean:			2.95370

Two-Way Anovar: Not significant.]

11. Would it be a good idea to set up a "laboratory" where students could practice, video tape, and see their speech before giving it in class?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	38.0%	30.6%	16.7%	3.7%	4.6%	6.5%	--%

	Variables	Categories	Means
[Group Mean Resp.:	1 (Treatment)	1 (Videotape)	1.86111
		2 (Audiotape)	2.36111
		3 (H. control)	2.55556
	2 (Instructor)	1 (Instructor #1)	2.16667
		2 (Instructor #2)	2.35185
Grand Mean:			2.25926

Two-way Anovar: Not significant.]

12. If such a lab existed, would you have used it two or more times this semester?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	28.7%	33.3%	16.7%	8.3%	11.1%	1.9%	--

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.00000
		2 (Audiotape)	2.72222
		3 (H. control)	2.63889
	2 (Instructor)	1 (Instructor #1)	2.42593
		2 (Instructor #2)	2.48148
Grand Mean:			2.45370

Two-Way Anovar: Not significant.]

13. Do you feel more confident to stand and speak to a group than you did at the beginning of the course?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	41.7%	37.0%	11.1%	4.6%	2.8%	1.9%	0.9%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.16667
		2 (Audiotape)	1.83333
		3 (H. control)	1.97222
	2 (Instructor)	1 (Instructor #1)	2.11111
		2 (Instructor #2)	1.87037
Grand Mean:			1.99074

Two-Way Anovar: Not significant.]

Section II. Answer only those parts which apply to you:*

Part A: Did you see video tape recordings of your 1st and 5th speeches in Roosevelt #108?

YES	NO
If so, answer the following questions:	If not, go on to the next part, p. 7.

1. Do you think this activity helped you improve your speaking?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	13.0%	26.9%	22.2%	16.7%	12.0%	4.6%	4.6%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	2.50000
		2 (Audiotape)	2.80556
		3 (H. control)	4.30556
	2 (Instructor)	1 (Instructor #1)	3.05556
		2 (Instructor #2)	3.35185
Grand Mean:			3.20370

Two-Way Anovar: $F_A = 9.789$ ($p < .01$, df = 2,102)

Tukey Post Hoc: Treatment 1 < Tr 3 ($p < .01$, df = 102)
Tr 2 < Tr 3 ($p < .01$, df = 102)]

2. Did it improve your poise and confidence?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	1.9%	22.2%	14.8%	24.1%	12.0%	22.2%	2.8%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	3.41667
		2 (Audiotape)	3.91667
		3 (H. control)	4.66667
	2 (Instructor)	1 (Instructor #1)	4.00000
		2 (Instructor #2)	4.00000
Grand Mean:			4.00000

---

*Responses of all subjects to this question, asked in Parts A, B, and C.

Two-Way Anovar: $F_A = 6.148$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)]

3. Did it help you in delivery?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	6.5%	34.3%	27.8%	11.1%	7.4%	10.2%	2.8%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	2.47222
		2 (Audiotape)	3.11111
		3 (H. control)	4.02778
	2 (Instructor)	1 (Instructor #1)	3.29630
		2 (Instructor #2)	3.11111
Grand Mean:			3.20370

Two-Way Anovar: $F_A = 11.138$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)
$A_2 < A_3$ ($p < .05$, df = 102)]

4. Did it help you in composition (organization, material, etc.)?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	0.9%	16.7%	17.6%	20.4%	18.5%	23.1%	2.8%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	4.13889
		2 (Audiotape)	4.00000
		3 (H. control)	4.44444
	2 (Instructor)	1 (Instructor #1)	4.18519
		2 (Instructor #2)	4.20370
Grand Mean:			4.19444

Two-Way Anovar: Not significant.]

5. Did it help you in language usage?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	6.5%	19.4%	21.3%	15.7%	15.7%	18.5%	2.8%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	3.58333
		2 (Audiotape)	3.30556
		3 (H. control)	4.55556
	2 (Instructor)	1 (Instructor #1)	3.77778
		2 (Instructor #2)	3.85185
Grand Mean:			3.81481

Two-Way Anovar: $F_A = 6.379$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .05$, df = 102)]

6. If you had not been required to view these tapes, would you have?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	18.5%	25.0%	21.3%	8.3%	13.0%	9.3%	4.6%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.44444
		2 (Audiotape)	2.97222
		3 (H. control)	4.13889
	2 (Instructor)	1 (Instructor #1)	3.29630
		2 (Instructor #2)	3.07407
Grand Mean:			3.18519

Two-Way Anovar: $F_A = 9.871$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)]
$A_2 < A_3$ ($p < .05$, df = 102)

7. Do you think this activity would be a good addition to this course?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	24.3%	31.8%	15.9%	14.0%	1.9%	9.3%	2.8%

	Variables	Categories	Means
Group Mean Resp.	1 (Treatment)	1 (Videotape)	1.91667
		2 (Audiotape)	2.33333
		3 (H. control)	3.97222
	2 (Instructor)	1 (Instructor #1)	2.66667
		2 (Instructor #2)	2.81481
Grand Mean:			2.74074

Two-Way Anovar: $F_A = 21.337$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)

$A_2 < A_3$ ($p < .01$, df = 102)]

8. After viewing did you find your appraisal of your performance differed much from your instructor's appraisal?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	1.9%	10.2%	14.8%	6.5%	10.2%	51.9%	4.6%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	5.02778
		2 (Audiotape)	4.91667
		3 (H. control)	4.66667
	2 (Instructor)	1 (Instructor #1)	4.85185
		2 (Instructor #2)	4.88889
Grand Mean:			4.87037

Two-Way Anovar: Not significant.]

9. Do you think the recording gave an accurate re-creation of how you really looked and sounded during these speeches?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

10. Did viewing your performances change your "self-image"?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

11. Had you thought you were better than the tape made you out to be?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

12. Do you think that the replay would be more helpful if it were done immediately after you gave the speech?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

13. When you gave your 1st speech, did you know it was being recorded?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	10.2%	18.5%	3.7%	0.9%	1.9%	40.7%	24.1%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	5.22222
		2 (Audiotape)	4.11111
		3 (H. control)	5.19444
	2 (Instructor)	1 (Instructor #1)	4.61111
		2 (Instructor #2)	5.07407
Grand Mean:			4.84259

Two-Way Anovar: Not significant.]

14. In rounds #1, #5, and #8, did the thought of speaking before a TV camera and mike increase your anxiety over speaking?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	4.6%	16.7%	13.0%	5.6%	10.2%	38.0%	12.0%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	4.77778
		2 (Audiotape)	4.44444
		3 (H. control)	4.63889
	2 (Instructor)	1 (Instructor #1)	4.57407
		2 (Instructor #2)	4.66667
Grand Mean:			4.62037

Two-Way Anovar: Not significant.]

15. Do you feel that you could have done a better job in these rounds if the camera and mike had not been present?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	1.9%	6.5%	9.3%	12.0%	13.9%	41.7%	14.8%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	5.50000
		2 (Audiotape)	4.91667
		3 (H. control)	5.00000
	2 (Instructor)	1 (Instructor #1)	5.07407
		2 (Instructor #2)	5.20370

Grand Mean: 5.13889

Two-Way Anovar: Not significant.]

16. Would it have been more distracting while you spoke to have an engineer in the class room running the TV camera?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	33.3%	33.3%	13.0%	4.6%	2.8%	11.1%	1.9%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	2.38889
		2 (Audiotape)	2.66667
		3 (H. control)	2.47222
	2 (Instructor)	1 (Instructor #1)	2.51852
		2 (Instructor #2)	2.50000

Grand Mean: 2.50926

Two-Way Anovar: Not significant.]

17. Did you find any aspect of the playback situation in Roosevelt distracting? (e.g. Personnel, equipment, room) (If so, circle)

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	8.3%	16.7%	3.7%	6.5%	4.6%	54.6%	5.6%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	3.91667
		2 (Audiotape)	5.08333
		3 (H. control)	5.05556
	2 (Instructor)	1 (Instructor #1)	4.96296
		2 (Instructor #2)	4.40741
Grand Mean:			4.68519

Two-Way Anovar: $F_A = 4.650$ ($p < .05$, df = 2,102)

Tukey Post Hoc: $A_1 > A_2$ ($p < .05$, df = 102)

$A_2 > A_3$ ($p < .05$, df = 102)]

18. Do you think that seeing the playback of your speeches made you more nervous and anxious about your speaking than you otherwise would have been?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	0.9%	2.8%	12.0%	8.3%	14.8%	54.6%	6.5%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	4.88889
		2 (Audiotape)	5.47222
		3 (H. control)	5.33333
	2 (Instructor)	1 (Instructor #1)	5.12963
		2 (Instructor #2)	5.33333
Grand Mean:			5.23148

Two-Way Anovar: Not significant.]

19. Was the playback experience improved by viewing your speech twice?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

20. Would it have helped to have a check-list to fill out as you watched your recording?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

21. Was it helpful to view another student's speech before seeing your own?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

22. Would it have helped to have your instructor's written critique while viewing?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

23. Did viewing your speeches motivate you to work more on speaking than you otherwise would have?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	10.2%	31.5%	25.0%	12.0%	8.3%	12.0%	0.9%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.58333
		2 (Audiotape)	2.97222
		3 (H. control)	3.94444
	2 (Instructor)	1 (Instructor #1)	3.25926
		2 (Instructor #2)	3.07407
Grand Mean:			3.16667

Two-Way Anovar: $F_A = 8.257$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)
$A_2 < A_3$ ($p < .05$, df = 102)]

24. Was viewing as helpful as your instructor's oral and written critique?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Total Resp.:	12.0%	22.2%	17.6%	7.4%	10.2%	24.1%	6.5%

	Variables	Categories	Means
Group Mean Resp.:	1(Treatment)	1 (Videotape)	2.80556
		2 (Audiotape)	3.83333
		3 (H. control)	4.75000
	2 (Instructor)	1 (Instructor #1)	3.77778
		2 (Instructor #2)	3.81481
Grand Mean:			3.79630

Two-Way Anovar: $F_A = 10.456$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_2$ ($p < .05$, df = 102)
$A_1 < A_3$ ($p < .01$, df = 102)]

25. Which of the following was most helpful for your speaking?

	Class Critiques	Instructor Critique	Viewing Tape	Writing Self-Evaluation
[Total Resp.:	20.4%	58.3%	16.7%	4.6%]

26. Do you think that viewing your tapes will have long range benefits for your speaking?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	10.2%	26.9%	13.0%	17.6%	14.8%	14.8%	2.8%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.75000
		2 (Audiotape)	3.41667
		3 (H. control)	4.50000
	2 (Instructor)	1 (Instructor #1)	3.55556
		2 (Instructor #2)	3.55556
Grand Mean:			3.55556

Two-Way Anovar: $F_A = 11.170$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)
$A_2 < A_3$ ($p < .05$, df = 102)]

27. Do you think you received as much "special attention" in this research project as students who engaged in other activities?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	6.5%	26.9%	18.5%	30.6%	6.5%	9.3%	1.9%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	3.41667
		2 (Audiotape)	2.83333
		3 (H. control)	3.91667
	2 (Instructor)	1 (Instructor #1)	3.35185
		2 (Instructor #2)	3.42593
Grand Mean:			3.38889

Two-Way Anovar: $F_A = 5.383$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_2 < A_3$ ($p < .01$, df = 102)]

Part B: Did you hear audio tape recordings of your 1st and 5th speeches in the Library's Audio-Visual Department?

YES	NO
If so, answer the following questions:	If not, go on to the next part, p. 10.

1. Do you think this activity helped you improve your speaking?

   Def. Yes — Yes — Prob. Yes — Undecided — Prob. No — No — Def. No

2. Did it improve your poise and confidence?

   Def. Yes — Yes — Prob. Yes — Undecided — Prob. No — No — Def. No

3. Did it help you in delivery?

   Def. Yes — Yes — Prob. Yes — Undecided — Prob. No — No — Def. No

4. Did it help you in composition (organization, material, etc.)?

   Def. Yes — Yes — Prob. Yes — Undecided — Prob. No — No — Def. No

5. Did it help you in language usage?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

6. If you had not been required to hear these tapes, would you have?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

7. Do you think this activity would be a good addition to this course?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

8. After hearing the tape did you find your appraisal of your performance differed much from your instructor's appraisal?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

9. Do you think the recording gave an accurate recreation of how you really sounded during these speeches?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

10. Did hearing your performance change your "self-image"?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

11. Had you thought you were better than the tape made you out to be?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

12. Do you think that the replay would be more helpful if it were done immediately after you gave the speech?

Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

13. When you gave your 1st speech, did you know it was being recorded?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

14. In rounds #1, #5, and #8, did the thought of speaking before a TV camera and mike increase your anxiety over speaking?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

15 Do you feel that you could have done a better job in these rounds if the camera and mike had not been present?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

16. Would it have been more distracting while you spoke to have an engineer in the class room running the TV camera?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

17. Did you find any aspect of the playback situation at Audio-Visual distracting? (e.g. personnel, equipment, room) (If so, circle)

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

18. Do you think that hearing the playback of your speeches made you more nervous and anxious about your speaking than you otherwise would have been?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

19. Was the playback experience improved by hearing your speech twice?

Def. Yes | Yes | Prob. Yes | Undecided | Prob. No | No | Def. No

20. Did you listen to your speech more than twice?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

21. Would it have helped to have a check-list to fill out as you heard your recording?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

22. Was it helpful to hear another student's speech before hearing your own?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

23. Would it have helped to have your instructor's written critique while listening?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

24. Did hearing your speeches motivate you to work more on speaking than you otherwise would have?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

25. Was listening to the replay as helpful as your instructor's oral and written critique?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

26. Which of the following was most helpful:

Class Critiques	Instructor Critique	Hearing Tape	Writing Self-Evaluation

27. Do you think that hearing your tapes will have long range benefits for your speaking?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

28. Do you think that you received as much "special attention" in this research project as students who engaged in other activities?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

Part C: Did you participate in the "Special Speech Improvement Activity" held in Roosevelt #108? (i.e. Viewing video taped performances of a student, Pres. Sponberg, and "Mark Twain")

YES	NO
If so, answer the following questions:	If not, go on to the next part.

1. Do you think this activity helped you improve your speaking?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

2. Did it improve your poise and confidence?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

3. Did it help you in delivery?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

4. Did it help you in composition (organization, material, etc.)?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

5. Did it help you in language usage?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

6. If you had not been required to view these tapes, would you have?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

7. Do you think this activity would be a good addition to this course?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

8. Did viewing these performances change your "self-image"?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

9. When you gave your 1st speech, did you know it was being recorded?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

10. In rounds #1, #5, and #8, did the thought of speaking before a TV camera and mike increase your anxiety over speaking?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

11. Do you feel that you could have done a better job in these rounds if the camera and mike had not been present?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

12. Would it have been more distracting while you spoke to have an engineer in the class room running the TV camera?

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

13. Did you find any aspect of the playback situation at Roosevelt distracting (e.g. Personnel, equipment, room) (If so, circle)

Def.		Prob.		Prob.		Def.
Yes	Yes	Yes	Undecided	No	No	No

14. Do you think that viewing these speeches made you more nervous and anxious about your speaking than you otherwise would have been?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

15. Did viewing these speeches motivate you to work more on speaking than you otherwise would have?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

16. Was viewing the speeches as helpful as your instructor's oral and written critique?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

17. Which of the following was most helpful for your speaking?

Class Critiques	Instructor Critique	Viewing "Special" Tape	Writing Self-Evaluation

18. Do you think that viewing this tape will have long range benefits for your speaking?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

19. Do you think you received as much "special attention" in this research project as students who engaged in other activities?

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

Part D: Were you in Section 2 (10:00, Mr. Mulac)?

Yes	No
If so, answer the following questions	If not, go on to the third section, p. 13.

1. Do you think that writing self-evaluations helped you improve your speaking?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

2. Did this activity improve your poise and confidence?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

3. Did it help you in delivery?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

4. Did it help you in composition (organization, material, etc.)?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

5. Did it help you in language usage?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

6. Do you think this activity would be a good addition to this course?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

7. Did it change your "self-image"?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

8. Which of the following was most helpful for your speaking?

   Class Critiques   Instructor Critique   Writing Self-Evaluation

9. Do you think you received as much "special attention" in this research project as students who engaged in other activities?

   Def. Yes   Yes   Prob. Yes   Undecided   Prob. No   No   Def. No

Section III. Everyone Answer These:

1. Do you think that students who were permitted to see video tape replays of their 1st and 5th speeches had an advantage in this course over those who didn't?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	22.2%	38.0%	25.9%	6.5%	1.9%	4.6%	0.9%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	2.38889
		2 (Audiotape)	2.58333
		3 (H. control)	2.38889
	2 (Instructor)	1 (Instructor #1)	2.46296
		2 (Instructor #2)	2.44444
Grand Mean:			2.45370

Two-Way Anovar: AB interaction = 3.400 ($p < .05$, df = 2,102)]

2. Do you think that students who were permitted to hear audio tape replays of their 1st and 5th speeches had an advantage over those who didn't?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	15.7%	37.0%	18.5%	11.1%	9.3%	6.5%	1.9%

	Variables	Categories	Means
Group Mean Resp.:	1 (Treatment)	1 (Videotape)	3.25000
		2 (Audiotape)	3.02778
		3 (H. control)	2.36111
	2 (Instructor)	1 (Instructor #1)	2.75926
		2 (Instructor #2)	3.00000
Grand Mean:			2.87963

Two-Way Anovar: F_A = 3.377 ($p < .05$, df = 2,102)
AB interaction = 5.426 ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 > A_3$ ($p < .05$, df = 102)]

3. Do you think that students who participated in the "Special Speech Improvement Activity" (in which they viewed and critiqued speeches of a student, Pres. Sponberg, and "Mark Twain") had an advantage over those who didn't?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Total Resp.:	8.3%	13.9%	22.2%	24.1%	16.7%	11.1%	3.7%

Group Mean Resp.:	Variables	Categories	Means
	1 (Treatment)	1 (Videotape)	3.38889
		2 (Audiotape)	3.33333
		3 (H. control)	4.52778
	2 (Instructor)	1 (Instructor #1)	3.79630
		2 (Instructor #2)	3.70370
Grand Mean:			3.75000

Two-Way Anovar: $F_A = 7.591$ ($p < .01$, df = 2,102)

Tukey Post Hoc: $A_1 < A_3$ ($p < .01$, df = 102)
$A_2 < A_3$ ($p < .01$, df = 102)]

4. Who do you believe is in charge of this research program?

[Total Resp.:		
17.6%	1.	Department of Speech
43.5%	2.	Mr. Webber (Coordinator of Instructional Broadcasting)
1.9%	3.	Mr. Mulac
7.4%	4.	Dr. Sattler (Chairman, Dept. of Speech)
26.9%	5.	Director of Office of Education, U. S. Dept. of H., E. & W.
2.8%]	6.	Mr. Schreiber

5. What do you think was the main purpose of this research project?

[Total Resp.:		
6.5%	1.	To study effects of anxiety producing stimulae on speakers.
16.7%	2.	To determine whether it is feasible to video tape and audio tape large numbers of student speeches.

4.6% 3. To determine whether students like to have their speeches recorded and played back.

5.6% 4. To determine the cost of a program of using different recording techniques in speech courses.

66.7%] 5. To determine whether showing a student his speeches helps him to improve later speaking.

THANK YOU FOR FILLING OUT THIS QUESTIONNAIRE!

(Is your name on the 1st page so you can receive credit?)

APPENDIX L*

QUESTIONNAIRE: JUDGING SESSIONS

Name:______________________________

Circle the group to which you belong:

Faculty Grad Student Undergrad Student

Based on your experiences in judging videotaped speeches by Fundamentals of Speech students please answer the following questions. Circle the response most in keeping with your honest feelings regarding each.

QUESTIONS:

1. Did you feel that the University of Wisconsin Speech Performance Scale provided a meaningful measure of "speech skill"?

	Definitely Yes	Yes	Probably Yes	Undecided	Probably No	No	Definitely No
Responses							
[Faculty:	--	1	2	--	--	1	1]
[Student:	--	2	1	--	--	1	--]

2. Did the individual factors (e.g. content, personality, voice, etc.) seem to be reasonable components of "speech skill"?

	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
Responses							
[Faculty:	1	1	--	1	--	2	--]
[Student:	--	1	1	1	--	1	--]

3. For each factor, was it possible for you to make seven discriminations? (i.e. 7 point scale reasonable?)

Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No

[omitted]

---

*Text and responses for the major experiment post-judging questionnaire.

4. Was it possible for you to make eleven discriminations per factor? (i.e. 11 point scale reasonable?)

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	1	2	1	--	1	--	--]
[Student:	--	2	2	--	--	--	--]

5. Which of the two scale lengths, 7 point or 11 point, permitted you to make more precise judgments? 7 point or 11 point.

Responses	7 point	11 point
[Faculty:	1	4]
[Student:	--	4]

6. Was the training session long enough for you to become familiar with the 6-factor rating scale?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	3	2	--	--	--	--	--]
[Student:	1	3	--	--	--	--	--]

7. Was the training session helpful in increasing your ability to use the scale with precision?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	3	--	1	1	--	--	--]
[Student:	1	3	--	--	--	--	--]

8. Did the lecture-discussion method used in the training session seem the appropriate method for training judges?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	3	1	1	--	--	--	--]
[Student:	1	2	1	--	--	--	--]

9. Was the technical quality of the video tape recordings so poor that it seemed to hinder the precision of your judgments?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	--	1	--	1	2	1]
[Student:	--	--	--	1	1	2	--]

10. To what extent was the picture and sound quality distracting?

Responses	Made Judging Impossible	Quite Distracting (Judging Possible)	Mildly Distracting	Not Distracting
[Faculty:	--	--	2	3]
[Student:	--	--	3	1]

11. Were you getting so tired and/or bored toward the end that your evaluations of the last speeches were less precise than earlier evaluations?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	--	1	--	4	--	--]
[Student:	--	--	1	2	1	--	--]

12. Do you think that this scale would be useful in speech research?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	2	2	--	1	--	--]
[Student:	--	3	--	--	1	--	--]

13. Do you feel that this scale would be useful in classroom teaching?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	--	1	3	--	--	--	1]
[Student:	--	2	1	--	1	--	--]

14. Would training in the use of this scale help inexperienced teachers understand what makes good speech performances?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	1	--	1	1	2	--	--]
[Student:	1	--	2	--	--	1	--]

15. Do you think this 6-factor scale is a better measure of speech skill than the single-factor scale of "General Effectiveness"?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	1	1	1	--	--	2	--]
[Student:	--	1	1	1	--	1	--]

16. Does this scale make for more precise evaluation of overt speech skill than the method used in the speech classes you teach or take?

Responses	Def. Yes	Yes	Prob. Yes	Undecided	Prob. No	No	Def. No
[Faculty:	1	--	1	1	--	1	1]
[Student:	--	1	2	--	--	1	--]

17. What parts of the scale, if any, did you find confusing?

[Responses: Occasional conflict in quality of sub-factors caused problems in giving one rating for that factor--2 judges.
Overlapping factors--2 judges.
"Personality" factor--2 judges.
"Intelligibility" factor--1 judge.
"Language" factor--1 judge.]

18. Do you have any suggestions for improving the training session?

[Responses: "No"--6.
"More of the range could be shown to anchor the scales--1 judge.
"Time"--1 judge.
Explain factor weighting system--1 judge.
Change "Personality" factor designation--1 judge.]

19. Do you have any suggestions for improving the use of this scale in speech research?

[Responses: "Content" too broad--1 judge.
Better conception of the factors--1 judge.
Better factor labels needed--1 judge.
Omit sub-factors--1 judge.
Taking notes--1 judge.]

20. Would judging 2-minute segments of these speeches permit a valid judgment?

[Responses: "Prob. Yes"--2 judges.
"Undecided"--2 judges.
"Prob. No"--1 judge.
"No"--1 judge.
"Prob. yes, except for content factor--2 judges.

APPENDIX M*

TABLE 14

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON TOTAL WEIGHTED$_{\text{Judges}}$ SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	2760.5625	107		
A (3 treatments)	56.6250	2	28.3125	(1.1363)[a]
B (2 instructors)	3.1250	1	3.1250	0.1254
AB	159.4375	2	79.7187	3.1996[b]
Subj. W. Groups	2541.3750	102	24.9154	
Within Subjects	936.2500	108		
C (2 measurements)	443.7500	1	443.7500	(97.0777)[a]
AC	10.6250	2	5.3125	(1.1622)[a]
BC	8.6875	1	8.6875	1.9005
ABC	6.9375	2	3.4687	0.7588
C X Subj.W. Groups	466.2500	102	4.7511	

[a] Planned comparison t tests were conducted in place of the F tests presented here within parentheses.

[b] $(p < .05, df = 2, 102)$

TABLE 15

RESULTS OF PLANNED COMPARISONS ON TOTAL WEIGHTED$_{\text{Judges}}$ SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.638
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.416
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	2.580[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	0.516
$C_2 \leq C_1$ (Post-test ≤ Pretest)	5.697[b]

[a] $(p < .01, df = 102)$

[b] $(p < .001, df = 102)$

---

*Results of three-way analyses of variance with repeated measures, and planned comparisons based on Price scale total weighted scores and factor scores for 216 speeches.

TABLE 16

RESULTS OF 3-WAY[1] ANALYSIS OF VARIANCE ON FACTOR #1, "CONTENT," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	294.0742	107		
A (3 treatments)	3.2500	2	1.6250	(0.5997)
B (2 instructors)	0.7969	1	0.7969	0.2941
AB	13.6484	2	6.8242	2.5185
Subj. W. Groups	276.3789	102	2.7096	
Within Subjects	210.0977	108		
C (2 measurements)	109.1016	1	109.1016	(115.3291)
AC	1.4180	2	0.7090	(0.7495)
BC	0.4844	1	0.4844	0.5120
ABC	2.6016	2	1.3008	1.3750
C X Subj.W. Groups	96.4922	102	0.9460	

TABLE 17

RESULTS OF PLANNED COMPARISONS ON FACTOR #1, "CONTENT," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.387
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.387
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	1.309
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	-0.174
$C_2 \leq C_1$ (Post-test ≤ Pretest)	12.396[a]

[a] ($p < .001$, df = 102)

TABLE 18

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON FACTOR #2, "INTELLIGIBILITY," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	441.0664	107		
A (3 treatments)	6.1875	2	3.0938	(0.7714)
B (2 instructors)	0.6484	1	0.6484	0.1617
AB	25.1641	2	12.5820	3.1373
Subj. W. Groups	409.0664	102	4.0105	
Within Subjects	76.4727	108		
C (2 measurements)	2.8125	1	2.8125	(3.9661)
AC	0.5234	2	0.2617	(0.3691)
BC	0.9141	1	0.9141	1.2890
ABC	-0.1094	2	-0.0547	-0.0771[a]
C X Subj.W. Groups	72.3320	102	0.7091	

[a] The result of rounding error.

TABLE 19

RESULTS OF PLANNED COMPARISONS ON FACTOR #2, "INTELLIGIBILITY," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.615
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.212
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	1.211
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	1.110
$C_2 \leq C_1$ (Post-test ≤ Pretest)	1.604

TABLE 20

MEANS AND STANDARD DEVIATIONS FOR INSTRUCTOR-REPEATED MEASURE GROUP INTERACTIONS ON FACTOR #3, "BODILY ACTION," SCORES

Measurement	Pretest (C_1)		Post-test (C_2)	
Statistic	Mean	S.D.	Mean	S.D.
$\text{Instr}_{\text{P.A.}}$ (B_1)	5.93	1.62	7.01	1.64
$\text{Instr}_{\text{Th.}}$ (B_2)	6.00	1.76	6.35	1.77
All Groups	5.96	1.69	6.68	1.69

TABLE 21

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON FACTOR #3, "BODILY ACTION," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	517.1211	107		
A (3 treatments)	17.2930	2	8.6465	(1.8559)
B (2 instructors)	4.6680	1	4.6680	1.0019
AB	19.9492	2	9.9746	2.1410
Subj. W. Groups	475.2109	102	4.6589	
Within Subjects	134.6680	108		
C (2 measurements)	27.6367	1	27.6367	(30.2034)
AC	5.8555	2	2.9277	(3.1996)
BC	7.1211	1	7.1211	7.7824[a]
ABC	0.7227	2	0.3613	0.3949
C X Subj.W. Groups	93.3320	102	0.9150	

[a] ($p < .01$, df = 1,102)

TABLE 22

RESULTS OF PLANNED COMPARISONS ON FACTOR #3, "BODILY ACTION," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.511
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.826
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	2.796[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	1.598
$C_2 \leq C_1$ (Post-test ≤ Pretest)	3.196[b]

[a] ($p < .005$, df = 102)

[b] ($p < .001$, df = 102)

TABLE 23

RESULTS OF 3-WAY[1] ANALYSIS OF VARIANCE ON FACTOR #4, "PERSONALITY," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	495.8750	107		
A (3 treatments)	8.6406	2	4.3203	(0.9553)
B (2 instructors)	0.2969	1	0.2969	0.0656
AB	25.6562	2	12.8281	2.8366
Subj. W. Groups	461.2812	102	4.5224	
Within Subjects	127.7227	108		
C (2 measurements)	27.9648	1	27.9648	(30.6197)
AC	3.2266	2	1.6133	(1.7664)
BC	2.7852	1	2.7852	3.0496
ABC	0.5898	2	0.2949	0.3229
C X Subj.W. Groups	93.1562	102	0.9133	

TABLE 24

RESULTS OF PLANNED COMPARISONS ON FACTOR #4, "PERSONALITY," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.439
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.519
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	2.132[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	1.111
$C_2 \leq C_1$ (Post-test ≤ Pretest)	3.198[b]

[a] ($p < .025$, df = 102)

[b] ($p < .001$, df = 102)

TABLE 25

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON FACTOR #5, "LANGUAGE," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	268.8398	107		
A (3 treatments)	3.3867	2	1.6934	(0.7126)
B (2 instructors)	0.0898	1	0.0898	0.0378
AB	22.9648	2	11.4824	4.8317
Subj. W. Groups	242.3984	102	2.3765	
Within Subjects	83.1094	108		
C (2 measurements)	21.0469	1	21.0469	(35.2631)
AC	0.7773	2	0.3887	(0.6512)
BC	0.0625	1	0.0625	0.1047
ABC	0.3438	2	0.1719	0.2880
C X Subj.W. Groups	60.8789	102	0.5969	

TABLE 26

RESULTS OF PLANNED COMPARISONS ON FACTOR #5, "LANGUAGE," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	0.661
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.110
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	2.089[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	0.660
$C_2 \leq C_1$ (Post-test ≤ Pretest)	3.463[b]

[a] ($p < .025$, df = 102)

[b] ($p < .001$, df = 102)

TABLE 27

RESULTS OF 3-WAY$_1$ ANALYSIS OF VARIANCE ON FACTOR #6, "VOICE," SCORES

Source of Variation	SS	DF	MS	F
Between Subjects	418.0312	107		
A (3 treatments)	12.3281	2	6.1641	(1.6092)
B (2 instructors)	0.0469	1	0.0469	0.0122
AB	14.9414	2	7.4707	1.9503
Subj. W.Groups	390.7148	102	3.8305	
Within Subjects	114.1641	108		
C (2 measurements)	38.4883	1	38.4883	(54.4306)
AC	1.5391	2	0.7695	(1.0883)
BC	2.0703	1	2.0703	2.9279
ABC	-0.0586	2	-0.0293	-0.0414[a]
C X Subj.W. Groups	72.1250	102	0.7071	

[a] The result of rounding error.

TABLE 28

RESULTS OF PLANNED COMPARISONS ON FACTOR #6, "VOICE," SCORES

Null Hypothesis	t
$A_1 \leq A_2$ (Videotape ≤ Audiotape)	1.084
$A_2 \leq A_3$ (Audiotape ≤ H. control)	0.000
$A_1C_2 \leq A_2C_2$ (Vid. post-test ≤ Aud. post-test)	3.330[a]
$A_2C_2 \leq A_3C_2$ (Aud. post-test ≤ H. control post-test)	0.202
$C_2 \leq C_1$ (Post-test ≤ Pretest)	4.238[b]

[a] ($p < .001$, df = 102)

[b] ($p < .001$, df = 102)

APPENDIX N*

TABLE 29

RESULTS OF 2-WAY ANALYSIS OF VARIANCE OF SUBJECT PRETEST PERFORMANCES ON TOTAL WEIGHTED$_{\text{Price}}$ SCORES

Source of Variation	DF	SS	MS	F
Total	107	1678.9448		
A (3 treatments)	2	16.3957	8.1978	0.5254
B (2 instructors)	1	0.0044	0.0044	0.0002
AB (treat.-instr.)	2	71.1824	35.5912	2.2812[a]
Within Groups	102	1591.3623	15.6015	

[a] ($p < .15$, df = 2, 102)

---

*2-way analysis of subject pre-test Price scale ratings, and 1-way analyses of course speech points and course grades assigned by each instructor.

TABLE 30

RESULTS OF 1-WAY ANALYSIS OF VARIANCE ON SUBJECT TOTAL SPEECH POINTS ASSIGNED BY INSTRUCTOR #1

Source of Variation	DF	SS	MS	F
Total	53	16781.4260		
Between Groups	2	887.3704	443.6852	1.4236
Within Groups (error)	51	15894.0556	311.6481	

TABLE 31

RESULTS OF 1-WAY ANALYSIS OF VARIANCE ON SUBJECT TOTAL SPEECH POINTS ASSIGNED BY INSTRUCTOR #2

Source of Variation	DF	SS	MS	F
Total	53	17086.8334		
Between Groups	2	1356.3332	678.1666	2.1986
Within Groups (error)	51	15730.5002	308.4411	

TABLE 32

RESULTS OF 1-WAY ANALYSIS OF VARIANCE ON SUBJECT TOTAL COURSE POINTS ASSIGNED BY INSTRUCTOR #1

Source of Variation	DF	SS	MS	F
Total	53	36788.5371		
Between Groups	2	2701.5926	1350.7963	2.0210
Within Groups (error)	51	34086.9445	668.3714	

TABLE 33

RESULTS OF 1-WAY ANALYSIS OF VARIANCE ON SUBJECT TOTAL COURSE POINTS ASSIGNED BY INSTRUCTOR #2

Source of Variation	DF	SS	MS	F
Total	53	55438.8149		
Between Groups	2	7202.4814	3601.2407	3.8075[a]
Within Groups (error)	51	48236.3335	945.8104	

[a] ($p < .05$, df = 2,51)

SELECTED BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

Books

Anastasi, Anne. Differential Psychology: Individual and Group Differences in Behavior. 3rd ed. New York: The MacMillan Company, 1958.

Barnlund, Dean C., and Haiman, Franklyn S. The Dynamics of Discussion. Boston: Houghton Mifflin Co., 1960.

Berlo, David K. The Process of Communication. New York: Holt, Rinehart, & Winston, 1960.

Bush, Robert N., McKnight, Philip, Clark, Richard J., Jr., Allen, Dwight W., Fortune, Jimmie C., and Cooper, James M. Micro Teaching: A Description. Palo Alto: Stanford University, 1968.

Butler, H. E. (trans.). The Institutio Oratoria of Quintilian. Cambridge: Harvard University Press, 1953.

Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963.

Cronbach, Lee J. Essentials of Psychological Testing. 2nd ed. revised. New York: Harper & Row, 1960.

Cureton, Edward E. "Validity," Educational Measurement. Edited by E. F. Lindquist. Washington, D. C.: American Council on Education, 1951.

Deese, James, and Hulse, Stewart. The Psychology of Learning. 3rd ed. New York: McGraw-Hill, 1967.

Edwards, Allen L. Experimental Design in Psychological Research. New York: Rinehart & Company, 1950.

Guilford, Joy P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1965.

________. Psychometric Methods. 2nd ed. revised. New York: McGraw-Hill, 1954.

Hays, William L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1958.

Osgood, Charles E., Suci, George J., and Tannenbaum, Percy H. The Measurement of Meaning. Illinois: University of Illinois Press, 1957.

Remmers, H. H. "Rating Methods in Research on Teaching," Handbook of Research on Teaching. Edited for the American Educational Research Association by Nathaniel L. Gage. Chicago: Rand McNally, 1967, 329-78.

Robinson, Karl F. Teaching Speech in the Secondary School. New York: Longmans, Green & Co., 1951.

Roethlisberger, Fritz, and Dickson, W. J. Management and the Worker. Cambridge: Harvard University Press, 1939.

Ross, Raymond. "The Wayne State University Program," The Student as Speaker and Listener. Edited by Elwood Murray and Arthur Solomon. Yellow Springs, Ohio: Antioch College and Jack Wolfram Foundation, 1966, 13-14.

Thompson, Wayne N. Quantitative Research in Public Address and Communication. New York: Random House, Inc., 1967.

Thorndike, Robert L. "Reliability," Educational Measurement. Edited by E. F. Lindquist. Washington, D. C.: American Council on Education, 1951.

Wiener, Norbert. The Human Use of Human Beings. Garden City, New York: Doubleday and Co., Inc., 1956.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., Inc., 1962.

Articles and Periodicals

Adams, Harlen M. "Magnetic Tape Voice Recording," Quarterly Journal of Speech, XXVII (October, 1941), 379-82.

Barker, Larry L., Kibler, Robert J., and Hunter, Eugenia C. "An Empirical Study of Overlap Rating Effects," Speech Teacher, XVII, No. 2 (March, 1968), 160-66.

Becker, Samuel L. "The Rating of Speeches: Scale Independence," Speech Monographs, XXIX (March, 1962), 38-44.

________, and Dallinger, Carl A. "The Effect of Instructional Methods Upon Achievement and Attitude in Communications Skills," Speech Monographs, XXVII (March, 1960), 70-76.

________, and Cronkhite, Gary L. "Reliability as a Function of Utilized Scale Steps," Speech Teacher, XIV (November, 1965), 291-93.

________, Bowers, John Waite, and Gronbeck, Bruce E. "Videotape in Teaching Discussion," Speech Teacher, XVII (March, 1968), 104-06.

Bettinghaus, Erwin P. "The Operation of Congruity in an Oral Communication Situation," Speech Monographs, XXVIII (August, 1961), 131-42.

Bilodeau, Edward A., and Bilodeau, Ina McD. "Motor-Skills Learning," Annual Review of Psychology, XXII (1961), 243-80.

________. "Variation of Temporal Intervals Among Critical Events in Five Studies of Knowledge of Results," Journal of Experimental Psychology, LV, No. 6 (1958), 603-12.

Bittner, Reign H., and Rundquist, Edward A. "The Rank-Comparison Rating Method," Journal of Applied Psychology, XXXIV (1950), 171-77.

Borchers, Gladys L. "Speech Without Work?" Quarterly Journal of Speech, XXI (June, 1935), 376-78.

Bowers, John Waite. "Training Speech Raters with Films," Speech Teacher, XIII (September, 1964), 228-31.

Broadrick, King, McIntyre, Charles J., and Moren, Richard. "TV Teacher's Report," Speech Teacher, XII (March, 1962), 152-57.

Brooks, Keith. "The Construction and Testing of a Forced Choice Scale for Measuring Speaking Achievement," Speech Monographs, XXIV (March, 1957), 65-73.

________, and Wulftange, Sr. I. Marie. "Listener Response to Oral Interpretation," Speech Monographs, XXXI (March, 1964), 73-79.

Brown, J. S. "A Proposed Program of Research on Psychological Feedback (Knowledge of Results) in the Performance of Psychomotor Tasks," Research Planning Conference on Perceptual and Motor Skills, AFHRRC Conf. Rept. 49-2 (U. S. Air Force, San Antonio, Tex., 1949), 81-87.

Bryan, Alice I., and Wilke, Walter H. "A Technique for Rating Public Speeches," Journal of Consulting Psychology, V (March, 1941), 80-91.

Caton, Chester F., and Feather, George K. "Teaching Speech with Television," NAEB Journal, XXIV (November-December, 1965), 24-26.

Clevenger, Theodore, and Cobin, Martin T. "An Experiment in Open Circuit Television Instruction in the Basic Course in Oral Interpretation," Speech Monographs, XXVI (June, 1959), 149-54.

Clevenger, Theodore, Jr. "Influence of Scale Complexity on the Reliability of Ratings of General Effectiveness in Public Speaking," Speech Monographs, XXXI (June, 1964), 153-56.

________. "Retest Reliabilities of Ten Scales of Public Speaking Performance," Central State Speech Journal, XIV (November 19, 1963), 285-91.

________. "Retest Reliability of Judgments of General Effectiveness in Public Speaking," Western Speech, XXVI (Fall, 1962), 216-22.

________, Clark, Margaret Leitner, and Lazier, Gilbert N. "Stability of Factor Structure in Smith's Semantic Differential for Theatre Concepts," Quarterly Journal of Speech, LIII (October, 1967), 241-47.

Cronkhite, Gary L. "Toward a Real Test of Dissonance Theory," Quarterly Journal of Speech, LII (April, 1966), 172-78.

Dedmon, Donald N., and Rayborn, David W. "Closed Circuit Television and the 'Required' First Course in Speech," Speech Teacher, XIV (November, 1965), 294-98.

Douglas, Jack. "The Measurement of Speech in the Classroom," Speech Teacher, VII (November, 1958), 309-19.

Drushal, J. Garber. "An Objective Analysis of Two Techniques of Teaching Delivery in Public Speaking," Quarterly Journal of Speech, XXV (October, 1939), 561-69.

Ebel, Robert L. "Estimation of the Reliability of Ratings," Psychometrika, XVI (December, 1951), 407-24.

Ehrensberger, Ray. "An Experimental Study of the Relative Effectiveness of Certain Forms of Emphasis in Public Speaking," Speech Monographs, XII (1945), 94-111.

Eisenstadt, Arthur. "As Others See Us," Speech Teacher, I (March, 1952), 121-24.

Faules, Donald. "The Relation of Communicator Skill to the Ability to Elicit and Interpret Feedback Under Four Conditions," Journal of Communication, XVII (December, 1967), 362-71.

Fotheringham, Wallace C. "A Technique for Measuring Speech Effectiveness in Public Speaking Classes," Speech Monographs, XXIII (March, 1956), 31-37.

Frandsen, Kenneth D. "Effects of Threat Appeals and Media of Transmission," Speech Monographs, XXX (June, 1963), 99-107.

Gibson, James W. "Using Videotape in the Training of Teachers," Speech Teacher, XVII (March, 1968), 107-09.

Gilkinson, Howard. "Experimental and Statistical Research in General Speech: I. Effects of Training and Correlates of Speech Skill," Quarterly Journal of Speech, XXX (February, 1944), 95-101.

________. "Experimental and Statistical Research in General Speech: II. Speakers, Speeches, and Audiences," Quarterly Journal of Speech, XXX (April, 1944), 180-86.

________, and Knower, Franklin H. "A Study of Standardized Personality Tests and Skill in Speech," Journal of Educational Psychology, XXXII (March, 1941), 161-75.

Gulley, Halbert E., and Berlo, David K. "Effect of Intercellular and Intracellular Speech Structure on Attitude Change and Learning," Speech Monographs, XXIII (August, 1956), 288-97.

Haines, Donald B. "Training for Culture-Contact and Interaction Skills," AMRL-TR-64-109. Wright-Patterson Air Force Base, Behavioral Sciences Laboratory, Air Force Systems Command, December 1964.

Haskitt, Harold O. "Using CCTV in Teaching Speech: The Moment of Truth," American Society for Engineering Education, LVI (November, 1965), 77.

Heinberg, Paul. "Relationships of Content and Delivery to General Effectiveness," *Speech Monographs*, XXX (June, 1963), 105-07.

Henderson, John. "Using Mirror TV to Teach Speaking," *NAEB Journal*, XXIII (November-December, 1964), 53-56.

Hirschfeld, Adeline Gittlen. "Videotape Recordings for Self-Analysis in the Speech Classroom," *Speech Teacher*, XVII (March, 1968), 116-18.

Holtzman, Paul D. "Speech Criticism and Evaluation as Communication," *Speech Teacher*, IX (January, 1960), 1-7.

Johnson, F. Craig, and Games, Paul A. "Effects on Learning of Adding 'Typical Questions' Responses to Television Lectures as a Type of Feedback," *Speech Monographs*, XXXII (November, 1963), 458-60.

________, Games, Paul A., Andersch, Elizabeth G., and Penson, Edward M. "An Investigation of Motion Picture Film and the Program Analyzer Feedback to Improve Television Teacher Training," *Speech Monographs*, XXX (November, 1963), 373-76.

Kanfer, Frederick H. "Verbal Rate, Content, and Adjustment Ratings in Experimentally Structured Interviews," *Journal of Abnormal and Social Psychology*, LVIII (1959), 305-11.

Kibler, Robert J., Barker, Larry L., and Enoch, Roy H. "The Development and Preliminary Assessment of a Set of Video-Taped Informative Speech Models," *Central States Speech Journal*, XVIII (November, 1967), 268-75.

Knower, Franklin H. "A Study of Speech Attitudes and Adjustments," *Speech Monographs*, V (1938), 130-203.

________. "What Is a Speech Test?" *Quarterly Journal of Speech*, XXX (February, 1944), 485-93.

Kramar, Edward J. J., and Lewis, Thomas R. "Comparison of Visual and Nonvisual Listening," *Journal of Communication*, I (November, 1951), 16-20.

Krasner, Leonard. "Studies of the Conditioning of Verbal Behavior," *Psychological Bulletin*, LV (1958), 148-70.

Kretsinger, Elwood A. "An Experimental Study of Gross Bodily Movement as an Index to Audience Interest," *Speech Monographs*, XIX (November, 1952), 244-48.

Laase, Leroy T. "An Evaluation of the Quality Rating System in Measuring Debate Achievement," Quarterly Journal of Speech, XXVIII (December, 1942), 424-30.

Leonardo, Manuel M. "An Experimental Study of a 'Teaching Machine' Application to Speech Improvement." Unpublished Ph.D. dissertation, University of Michigan, 1963. (Abstract: Speech Monographs, XXXI [August, 1964], 235.)

Mabie, E. C. "The Responses of Theater Audiences, Experimental Studies," Speech Monographs, XIX (November, 1952), 236-43.

McIntyre, Charles J. "Applying Learning Theory to Televised Instruction," NAEB Journal, XXIV (November-December, 1965), 54-70.

Miller, Gerald R. "Agreement and the Grounds for It: Persistent Problems in Speech Rating," Speech Teacher, XIII (November, 1964), 257-61.

________, Zavos, Harry, Vlandis, John W., and Rosenbaum, Milton E. "The Effect of Differential Reward on Speech Patterns," Speech Monographs, XXVIII (March, 1961), 9-15.

Miyamoto, S. Frank, Crowell, Laura, and Katcher, Allan. "Self-Concepts of Communicative Skill Among Beginning Speech Students," Speech Monographs, XXIII (March, 1956), 66-74.

Monroe, Alan H. "The Measurement and Analysis of Audience Reaction to Student Speakers--Studies in Attitude Changes," Studies in Higher Education, XXXII (December, 1937), 7-80.

________, Remmers, H. H., and Venemann-Lyle, Elizabeth. "Measuring the Effectiveness of Public Speech in a Beginning Course," Studies in Higher Education, XXIX (September, 1936), 5-29.

Moore, Floyd Jack, Chernell, Eugene, and West, Maxwell J. "'O Wad Some Power the Giftie Gie Us': Television as a Therapeutic Tool," Psychiatric Spectator, I (November, 1964), 14-15.

Moore, Wilbur E. "Factors Related to Achievement and Improvement in Public Speaking," Quarterly Journal of Speech, XXIX (April, 1943), 213-17.

Moss, C. S. "Current and Projected Status of Semantic Differential Research," Psychological Record, X (1960), 47-54.

Nelson, Harold. "The Use of Closed-Circuit TV in Teaching the Basic Speech Course," Speech Teacher, VII (January, 1958), 1-5.

Nelson, Harold E. "The Relative Contribution to Learning of Video and Audio Elements in Films," Speech Monographs, XVIII (March, 1951), 70-73.

________. "Videotaping the Speech Course," Speech Teacher, XVII (March, 1968), 101-03.

Norvelle, Lee. "Development and Application of a Method for Measuring the Effectiveness of Instruction in a Basic Speech Course," Speech Monographs, I (1934), 41-65.

Nystrom, Clarence L., and Leaf, Robert. "The Recording Machine as a Teaching Device," Quarterly Journal of Speech, XXV (October, 1939), 433-38.

Ochs, Donovan J. "Videotape in Teaching Advanced Public Speaking," Speech Teacher, XVII (March, 1968), 110-12.

Oliver, Robert T. "The Eternal (and Infernal) Problem of Grades," Speech Teacher, IX (January, 1960), 8-11.

Osband, Helen. "The Giftie Gie Us," Southern Speech Journal, III (November, 1937), 19-22.

Pelsma, J. R. "Standardization of Grades in Public Speaking," Quarterly Journal of Public Speaking, I (October, 1915), 266-71.

Phelps, Waldo, and Hempen, Claude. "Beginning Speeches and Closed Circuit TV," Western Speech Journal, XXIV (Winter, 1960), 33-36.

Popham, W. James. "Instructional Video Tapes in Teacher Education," AV Communication Review, XIV (Fall, 1966), 371-76.

Reynolds, R. V. E. "Videotape in Teaching Speech in a Small College," Speech Teacher, XVII (March, 1968), 113-15.

Richardson, Don, and Waal, Carla. "Increasing the Reliability of Judgments of Acting Performance," Quarterly Journal of Speech, LII (December, 1966), 378-82.

Saltzman, S. W. "TV Records Student Performances at University of Illinois," NAEB Journal, XXIII (March-April, 1964), 3-6.

Sawyer, Thomas Jr. "A Grading System for Speech Classes," Speech Teacher, IX (January, 1960), 12-15.

Scheidel, Thomas M. "Sex and Persuasibility," Speech Monographs, XXX (November, 1963), 353-58.

Simmons, I. F. "The Nature and Use of Audio-Visual Aids in Speech Instruction," Southern Speech Journal, IX (November, 1943), 215-19.

Simon, Clarence T. "Speech as a Science," Quarterly Journal of Speech, XXXVI (October, 1951), 292.

Smith, Raymond G. "Validation of a Semantic-Differential," Speech Monographs, XXX (March, 1963), 50-55.

________. "Development of a Semantic Differential for Use with Speech Related Concepts," Speech Monographs, XXVI (November, 1959), 263-72.

Smode, A. F. "Learning and Performance in a Tracking Task under Two Levels of Achievement Information Feedback," Journal of Experimental Psychology, LVI (October, 1958), 297-304.

Spitzer, Robert S., Jackson, Don D., and Satir, Virginia. "A New Technique for Teaching Conjoint Family Therapy," Psychiatric Spectator, I (1964), 13-14.

Stanmyre, R. William. "The Syracuse Film Experiments," Quarterly Journal of Speech, XXXVI (February, 1950), 57-60.

Stevens, Wilmer E. "A Rating Scale for Public Speakers," Quarterly Journal of Speech, XIV (April, 1928), 223-32.

Tannenbaum, P. H. "The Effect of Background Music on Interpretation of Stage and Television Drama," Audio-Visual Communications Review, IV (Spring, 1956), 92-101.

Tenney, William H. "The Measurement of Speech Recorded on Film," Quarterly Journal of Speech, XXXI (October, 1945), 370-73.

Thompson, Wayne. "Is There a Yardstick for Measuring Speaking Skill?" Quarterly Journal of Speech, XXIX (February, 1943), 87-91.

Thompson, Wayne N. "A Study of the Characteristics of Student Raters of Public Speaking Performances," Speech Monographs, XIII (1946), 45-53.

________. "An Experimental Study of the Accuracy of Typical Speech Rating Techniques," Speech Monographs, XI (1944), 65-79.

Utzinger, Vernon A. "The Use of the Moving Picture Machine and the Recording Instrument in Teaching Speech," Quarterly Journal of Speech, XXV (February, 1939), 9-12.

Vlandis, John W. "Variation in the Verbal Behavior of a Speaker as a Function of Varied Reinforcing Conditions," Speech Monographs, XXXI (June, 1964), 116-19.

Watkins, Dwight E. "An Apparatus for Recording Speeches," Quarterly Journal of Speech, X (1924), 253-58.

Wiksell, Wesley. "New Methods of Evaluating Instruction and Student Achievement in a Speech Class," Speech Teacher, IX (January, 1960), 16-19.

Wilke, Walter H. "A Speech Profile," Quarterly Journal of Speech, XXVI (December, 1940), 625-30.

Williamson, Arleigh. "Two Years' Experience with Recording Equipment," Quarterly Journal of Speech, XXI (October, 1935), 195-216.

Woodward, Howard S. "Measurement and Analysis of Audience Opinion," Quarterly Journal of Speech," XIV (February, 1928), 94-111.

Unpublished Works

Allen, Dwight W., and Fortune, Jimmie C. "An Analysis of Micro-Teaching: A New Procedure in Teacher Education." Unpublished paper, Stanford University, February 8, 1965. (Mimeographed.)

Andersen, Kenneth E. "An Experimental Study of the Artistic and Nonartistic Ethos in Persuasion." Unpublished Ph.D. dissertation, University of Wisconsin, 1961.

Becker, Samuel L. "Rating Scales." Unpublished chapter, draft, 1967. (Dittoed.)

———————. Comments at a meeting of the Behavioral Science Interest Group at S.A.A. Convention, Los Angeles, Dec. 28, 1967.

Clevenger, Theodore, Jr. "A Regression Study of the Price Rating Scales for Public Speaking with a Diversified Group of Student Judges." Unpublished report, University of Wisconsin Communication and Public Address Experimental Laboratory, No. JR-003, February 25, 1963.

Frandsen, K. D., Larson, C. E., and Knapp, M. L. "Some Effects of Simultaneous and Sequential Video-Tape Playback and Instructor Comments: A Preliminary Report." Unpublished paper presented at the Conference of the Directors of the Basic Course of the Midwestern Universities, Lawrence, Kansas, February 18, 1967.

Hirschfeld, Adeline Gittlen. "Utilization of Video Taped Speeches for Self Analysis in Fundamentals of Speech Course." Unpublished Ph.D. dissertation, Wayne State University, Detroit, 1966.

Holladay, Howard. "A Pilot Study of the Effects of Simultaneous, Immediate, and Delayed Videotape Playback on Self-Perception of Delivery Skills." Unpublished paper, University of Southern Califormia, Los Angeles, 1968.

Livingston, Harold M., and Doler, Thurston E. "Effectiveness of Video-Taped Feed-Back versus Conventional Oral Critique in Teaching the Basic College Speech Course." Unpublished paper, Oregon State University, Corvallis, 1964. (Mimeographed.)

Moser, Thomas F. "An Experimental Study of the Effects of Verbal Reinforcement on Small Group Discussion." Unpublished Ph.D. dissertation, University of Michigan, 1965.

Price, William K. "The University of Wisconsin Speech Attainment Test." Unpublished Ph.D. dissertation, University of Wisconsin, 1964.

Smith, Calvin N. "Evaluating Classroom Discussion." Unpublished paper presented at the Central States Speech Convention, Chicago, April 5, 1968.

Interviews

Bowers, John Waite. Telephone interview, October 20, 1967.

Becker, Samuel L. Telephone interview, March 4, 1968.

McCroskey, James C., and Lashbrook, William B. "Video Tape as a Self-Confrontation and Critical Device in the Teaching of Public Speaking." (Incomplete, unpublished research.) Telephone interview, August 1, 1968.

Strauss, Robert. Telephone interview, August 16, 1968.